Spectral Statistics of Erdős-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues

Spectral Statistics of Erdős-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues

László Erdős[1*] Antti Knowles[2†] Horng-Tzer Yau[2‡] Jun Yin[2§]

Institute of Mathematics, University of Munich,
Theresienstrasse 39, D-80333 Munich, Germany
lerdos@math.lmu.de [1]

Department of Mathematics, Harvard University
Cambridge MA 02138, USA
knowles@math.harvard.edu *htyau@math.harvard.edu* *jyin@math.harvard.edu* [2]

July 27, 2012

We consider the ensemble of adjacency matrices of Erdős-Rényi random graphs, i.e. graphs on N vertices where every edge is chosen independently and with probability $p \equiv p(N)$. We rescale the matrix so that its bulk eigenvalues are of order one. Under the assumption $pN \gg N^{2/3}$, we prove the universality of eigenvalue distributions both in the bulk and at the edge of the spectrum. More precisely, we prove (1) that the eigenvalue spacing of the Erdős-Rényi graph in the bulk of the spectrum has the same distribution as that of the Gaussian orthogonal ensemble; and (2) that the second largest eigenvalue of the Erdős-Rényi graph has the same distribution as the largest eigenvalue of the Gaussian orthogonal ensemble. As an application of our method, we prove the bulk universality of generalized Wigner matrices under the assumption that the matrix entries have at least $4 + \varepsilon$ moments.

AMS Subject Classification (2010): 15B52, 82B44

Keywords: Erdős-Rényi graphs, universality, Dyson Brownian motion.

[*]Partially supported by SFB-TR 12 Grant of the German Research Council
[†]Partially supported by NSF grant DMS-0757425
[‡]Partially supported by NSF grants DMS-0757425, 0804279
[§]Partially supported by NSF grant DMS-1001655

1. Introduction

The Erdős-Rényi ensemble [20, 21] is a law of a random graph on N vertices, in which each edge is chosen independently with probability $p \equiv p(N)$. The corresponding adjacency matrix is called the Erdős-Rényi matrix. Since each row and column has typically pN nonzero entries, the matrix is sparse as long as $p \ll 1$. We shall refer to pN as the sparseness parameter of the matrix. In the companion paper [11], we established the local semicircle law for the Erdős-Rényi matrix for $pN \geqslant (\log N)^C$, i.e. we showed that, assuming $pN \geqslant (\log N)^C$, the eigenvalue density is given by the Wigner semicircle law in any spectral window containing on average at least $(\log N)^{C'}$ eigenvalues. In this paper, we use this result to prove both the bulk and edge universalities for the Erdős-Rényi matrix under the restriction that the sparseness parameter satisfies

$$pN \gg N^{2/3}. \tag{1.1}$$

More precisely, assuming that p satisfies (1.1), we prove that the eigenvalue spacing of the Erdős-Rényi graph in the bulk of the spectrum has the same distribution as that of the Gaussian orthogonal ensemble (GOE). In order to outline the statement of the edge universality for the Erdős-Rényi graph, we observe that, since the matrix elements of the Erdős-Rényi ensemble are either 0 or 1, they do not satisfy the mean zero condition which typically appears in the random matrix literature. In particular, the largest eigenvalue of the Erdős-Rényi matrix is very large and lies far away from the rest of the spectrum. We normalize the Erdős-Rényi matrix so that the bulk of its spectrum lies in the interval $[-2, 2]$. By the edge universality of the Erdős-Rényi ensemble, we therefore mean that its second largest eigenvalue has the same distribution as the largest eigenvalue of the GOE, which is the well-known Tracy-Widom distribution. We prove the edge universality under the assumption (1.1).

Neglecting the mean zero condition, the Erdős-Rényi matrix becomes a Wigner random matrix with a Bernoulli distribution when $0 < p < 1$ is a constant independent of N. Thus for $p \ll 1$ we can view the Erdős-Rényi matrix, up to a shift in the expectation of the matrix entries, as a singular Wigner matrix for which the probability distributions of the matrix elements are highly concentrated at zero. Indeed, the probability for a single entry to be zero is $1 - p$. Alternatively, we can express the singular nature of the Erdős-Rényi ensemble by the fact that the k-th moment of a matrix entry is bounded by

$$N^{-1}(pN)^{-(k-2)/2}\,. \tag{1.2}$$

For $p \ll 1$ this decay in k is much slower than in the case of Wigner matrices.

There has been spectacular progress in the understanding of the universality of eigenvalue distributions for invariant random matrix ensembles [5, 7, 8, 27, 28]. The Wigner and Erdős-Rényi matrices are not invariant ensembles, however. The moment method [31, 33, 32] is a powerful means for establishing edge universality. In the context of sparse matrices, it was applied in [32] to prove edge universality for the zero mean version of the d-regular graph, where the matrix entries take on the values -1 and 1 instead of 0 and 1. The need for this restriction can be ascribed to the two following facts. First, the moment method is suitable for treating the largest and smallest eigenvalues. But in the case of the Erdős-Rényi matrix, it is the second largest eigenvalue, not the largest one, which behaves like the largest eigenvalue of the GOE. Second, the modification of the moment method to matrices with non-symmetric distributions poses a serious technical challenge.

A general approach to proving the universality of Wigner matrices was recently developed in the series of papers [12, 13, 14, 15, 16, 17, 18, 19]. In this paper, we further extend this method to cover sparse matrices such as the Erdős-Rényi matrix in the range (1.1). Our approach is based on the following three ingredients. (1) A local semicircle law – a precise estimate of the local eigenvalue density down to energy scales containing around $(\log N)^C$ eigenvalues. (2) Establishing universality of the eigenvalue distribution of Gaussian divisible ensembles, via an estimate on the rate of decay to local equilibrium of the Dyson Brownian motion [9]. (3) A density argument which shows that for any probability distribution of the matrix entries there exists a Gaussian divisible distribution such that the two associated Wigner ensembles have identical local eigenvalue statistics down to the scale $1/N$. In the case of Wigner matrices, the edge universality can also be obtained by a modification of (1) and (3) [19]. The class of ensembles to which this method applies is extremely general. So far it includes all (generalized) Wigner matrices under the sole assumption that the distributions of the matrix elements have a uniform subexponential decay. In this paper we extend this method to the Erdős-Rényi matrix, which in fact represents a generalization in two unrelated directions: (a) the law of the matrix entries is much more singular, and (b) the matrix elements have nonzero mean.

As an application of the local semicircle law for sparse matrices proved in [11], we also prove the bulk universality for generalized Wigner matrices under the sole assumption that the matrix entries have $4+\varepsilon$ moments. This relaxes the subexponential decay condition on the tail of the distributions assumed in [17, 18, 19]. Moreover, we prove the edge universality of Wigner matrices under the assumption that the matrix entries have $12+\varepsilon$ moments. These results on Wigner matrices are stated and proved in Section 7 below. We note that in [3] it was proved that the distributions of the largest eigenvalues are Poisson if the entries have at most $4-\varepsilon$ moments. Numerical results [4] predict that the existence of four moments corresponds to a sharp transition point, where the transition is from the Poisson process to the determinantal point process with Airy kernel.

We remark that the bulk universality for Hermitian Wigner matrices was also obtained in [34], partly by using the result of [22] and the local semicircle law from Step (1). For real symmetric Wigner matrices, the bulk universality in [34] requires that the first four moments of every matrix element coincide with those of the standard Gaussian random variable. In particular, this restriction rules out the real Bernoulli Wigner matrices, which may be regarded as the simplest kind of an Erdős-Rényi matrix (again neglecting additional difficulties arising from the nonzero mean of the entries).

As a first step in our general strategy to prove universality, we proved, in the companion paper [11], a local semicircle law stating that the eigenvalue distribution of the Erdős-Rényi ensemble in any spectral window which on average contains at least $(\log N)^C$ eigenvalues is given by the Wigner semicircle law. As a corollary, we proved that the eigenvalue locations are equal to those predicted by the semicircle law, up to an error of order $(pN)^{-1}$. The second step of the strategy outlined above for Wigner matrices is to estimate the local relaxation time of the Dyson Brownian motion [15, 16]. This is achieved by constructing a pseudo-equilibrium measure and estimating the global relaxation time to this measure. For models with nonzero mean, such as the Erdős-Rényi matrix, the largest eigenvalue is located very far from its equilibrium position, and moves rapidly under the Dyson Brownian motion. Hence a uniform approach to equilibrium is impossible. We overcome this problem by integrating out the largest eigenvalue from the joint probability distribution of the eigenvalues, and consider the flow of the marginal distribution of the remaining $N-1$ eigenvalues. This enables us to establish bulk universality for sparse matrices with nonzero mean under the restriction (1.1). This approach trivially also applies to Wigner matrices whose entries have nonzero mean.

Since the eigenvalue locations are only established with accuracy $(pN)^{-1}$, the local relaxation time for the Dyson Brownian motion with the initial data given by the Erdős-Rényi ensemble is only shown to be less than $1/(p^2N) \gg 1/N$. For Wigner ensembles, it was proved in [19] that the local relaxation time is of order $1/N$. Moreover, the slow decay of the third moment of the Erdős-Rényi matrix entries, as given in (1.2), makes the approximation in Step (3) above less effective. These two effects impose the restriction (1.1) in our proof of bulk universality. At the end of Section 2 we give a more detailed account of how this restriction arises. The reason for the same restriction's being needed for the edge universality is different; see Section 6.3. We note, however, that both the bulk and edge universalities are expected to hold without this restriction, as long as the graphs are not too sparse in the sense that $pN \gg \log N$; for d-regular graphs this condition is conjectured to be the weaker $pN \gg 1$ [30]. A discussion of related problems on d-regular graphs can be found in [26].

Acknowledgement. We thank P. Sarnak for bringing the problem of universality of sparse matrices to our attention.

2. Definitions and results

We begin this section by introducing a class of $N \times N$ sparse random matrices $A \equiv A_N$. Here N is a large parameter. (Throughout the following we shall often refrain from explicitly indicating N-dependence.)

The motivating example is the *Erdős-Rényi matrix*, or the adjacency matrix of the *Erdős-Rényi random graph*. Its entries are independent (up to the constraint that the matrix be symmetric), and equal to 1 with probability p and 0 with probability $1-p$. For our purposes it is convenient to replace p with the new parameter $q \equiv q(N)$, defined through $p = q^2/N$. Moreover, we rescale the matrix in such a way that its bulk eigenvalues typically lie in an interval of size of order one.

Thus we are led to the following definition. Let $A = (a_{ij})$ be the symmetric $N \times N$ matrix whose entries a_{ij} are independent (up to the symmetry constraint $a_{ij} = a_{ji}$) and each element is distributed according to

$$a_{ij} \;=\; \frac{\gamma}{q}\begin{cases} 1 & \text{with probability } \frac{q^2}{N} \\ 0 & \text{with probability } 1 - \frac{q^2}{N}\,. \end{cases} \tag{2.1}$$

Here $\gamma := (1 - q^2/N)^{-1/2}$ is a scaling introduced for convenience. The parameter $q \leqslant N^{1/2}$ expresses the sparseness of the matrix; it may depend on N. Since A typically has $q^2 N$ nonvanishing entries, we find that if $q \ll N^{1/2}$ then the matrix is sparse.

We extract the mean of each matrix entry and write

$$A \;=\; H + \gamma q \, |\mathbf{e}\rangle\langle\mathbf{e}| \,,$$

where the entries of H (given by $h_{ij} = a_{ij} - \gamma q/N$) have mean zero, and we defined the vector

$$\mathbf{e} \;\equiv\; \mathbf{e}_N \;:=\; \frac{1}{\sqrt{N}}(1, \dots, 1)^T \,. \tag{2.2}$$

Here we use the notation $|\mathbf{e}\rangle\langle\mathbf{e}|$ to denote the orthogonal projection onto $\mathbf{e}$, i.e. $(|\mathbf{e}\rangle\langle\mathbf{e}|)_{ij} := N^{-1}$.

One readily finds that the matrix elements of H satisfy the moment bounds

$$\mathbb{E} h_{ij}^2 \;=\; \frac{1}{N}\,, \qquad \mathbb{E}\big|h_{ij}\big|^p \;\leqslant\; \frac{1}{N q^{p-2}}\,, \tag{2.3}$$

where $p \geqslant 2$.

More generally, we consider the following class of random matrices with non-centred entries characterized by two parameters q and f, which may be N-dependent. The parameter q expresses how singular the distribution of h_{ij} is; in particular, it expresses the sparseness of A for the special case (2.1). The parameter f determines the nonzero expectation value of the matrix elements.

Definition 2.1 (H). *We consider $N \times N$ random matrices $H = (h_{ij})$ whose entries are real and independent up to the symmetry constraint $h_{ij} = h_{ji}$. We assume that the elements of H satisfy the moment conditions*

$$\mathbb{E}h_{ij} \;=\; 0\,, \qquad \mathbb{E}|h_{ij}|^2 \;=\; \frac{1}{N}\,, \qquad \mathbb{E}|h_{ij}|^p \;\leqslant\; \frac{C^p}{Nq^{p-2}} \tag{2.4}$$

for $1 \leqslant i,j \leqslant N$ and $2 \leqslant p \leqslant (\log N)^{10 \log\log N}$, where C is a positive constant. Here $q \equiv q(N)$ satisfies

$$(\log N)^{15 \log\log N} \;\leqslant\; q \;\leqslant\; CN^{1/2} \tag{2.5}$$

for some positive constant C.

Definition 2.2 (A). *Let H satisfy Definition 2.1. Define the matrix $A = (a_{ij})$ through*

$$A \;:=\; H + f|\mathbf{e}\rangle\langle\mathbf{e}|\,, \tag{2.6}$$

where $f \equiv f(N)$ is a deterministic number that satisfies

$$1 + \varepsilon_0 \;\leqslant\; f \;\leqslant\; N^C\,, \tag{2.7}$$

for some constants $\varepsilon_0 > 0$ and C.

Remark 2.3. For definiteness, and bearing the Erdős-Rényi matrix in mind, we restrict ourselves to real symmetric matrices satisfying Definition 2.2. However, our proof applies equally to complex Hermitian sparse matrices.

Remark 2.4. As observed in [11], Remark 2.5, we may take H to be a Wigner matrix whose entries have subexponential decay $\mathbb{E}|h_{ij}|^p \leqslant (Cp)^{\theta p} N^{-p/2}$ by choosing $q = N^{1/2}(\log N)^{-5\theta \log\log N}$.

We shall use C and c to denote generic positive constants which may only depend on the constants in assumptions such as (2.4). Typically, C denotes a large constant and c a small constant. Note that the fundamental large parameter of our model is N, and the notations $\gg, \ll, O(\cdot), o(\cdot)$ always refer to the limit $N \to \infty$. Here $a \ll b$ means $a = o(b)$. We write $a \sim b$ for $C^{-1}a \leqslant b \leqslant Ca$.

After these preparations, we may now state our results. They concern the distribution of the eigenvalues of A, which we order in a nondecreasing fashion and denote by $\mu_1 \leqslant \cdots \leqslant \mu_N$. We shall only consider the distribution of the $N-1$ first eigenvalues $\mu_1, \dots, \mu_{N-1}$. The largest eigenvalue μ_N lies far removed from the others, and its distribution is known to be normal with mean $f + f^{-1}$ and variance $N^{-1/2}$; see [11], Theorem 6.2, for more details.

First, we establish the bulk universality of eigenvalue correlations. Let $p(\mu_1, \dots, \mu_N)$ be the probability density[1] of the ordered eigenvalues $\mu_1 \leqslant \cdots \leqslant \mu_N$ of A. Introduce the marginal density

$$p_N^{(N-1)}(\mu_1, \dots, \mu_{N-1}) \;:=\; \frac{1}{(N-1)!} \sum_{\sigma \in S_{N-1}} \int \mathrm{d}\mu_N \, p(\mu_{\sigma(1)}, \dots, \mu_{\sigma(N-1)}, \mu_N)\,.$$

In other words, $p_N^{(N-1)}$ is the symmetrized probability density of the first $N-1$ eigenvalues of H. For $n \leqslant N-1$ we define the n-point correlation function (marginal) through

$$p_N^{(n)}(\mu_1, \dots, \mu_n) \;:=\; \int \mathrm{d}\mu_{n+1} \cdots \mathrm{d}\mu_{N-1} \, p_N^{(N-1)}(\mu_1, \dots, \mu_{N-1})\,. \tag{2.8}$$

Similarly, we denote by $p_{\mathrm{GOE},N}^{(n)}$ the n-point correlation function of the symmetrized eigenvalue density of an $N \times N$ GOE matrix.

[1]Note that we use the density of the law of the eigenvalue density for simplicity of notation, but our results remain valid when no such density exists.

Theorem 2.5 (Bulk universality). *Suppose that A satisfies Definition 2.2 with $q \geqslant N^{\phi}$ for some ϕ satisfying $0 < \phi \leqslant 1/2$, and that f additionally satisfies $f \leqslant C N^{1/2}$ for some $C > 0$. Let $\beta > 0$ and assume that*

$$\phi \; > \; \frac{1}{3} + \frac{\beta}{6} \,. \tag{2.9}$$

Let $E \in (-2, 2)$ and take a sequence (b_N) satisfying $N^{\varepsilon - \beta} \leqslant b_N \leqslant \lvert \lvert E \rvert - 2 \rvert / 2$ for some $\varepsilon > 0$. Let $n \in \mathbb{N}$ and $O : \mathbb{R}^n \to \mathbb{R}$ be compactly supported and continuous. Then

$$\lim_{N \to \infty} \int_{E - b_N}^{E + b_N} \frac{\mathrm{d}E'}{2 b_N} \int \mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_n \, O(\alpha_1, \dots, \alpha_n) \\ \times \frac{1}{\varrho_{sc}(E)^n} \big(p_N^{(n)} - p_{\mathrm{GOE},N}^{(n)} \big) \bigg(E' + \frac{\alpha_1}{N \varrho_{sc}(E)}, \dots, E' + \frac{\alpha_n}{N \varrho_{sc}(E)} \bigg) \; = \; 0 \,,$$

where we abbreviated

$$\varrho_{sc}(E) \; := \; \frac{1}{2\pi} \sqrt{[4 - E^2]_+} \tag{2.10}$$

for the density of the semicircle law.

Remark 2.6. Theorem 2.5 implies bulk universality for sparse matrices provided that $1/3 < \phi \leqslant 1/2$. See the end of this section for an account on the origin of the condition (2.9).

We also prove the universality of the extreme eigenvalues.

Theorem 2.7 (Edge universality). *Suppose that A satisfies Definition 2.2 with $q \geqslant N^{\phi}$ for some ϕ satisfying $1/3 < \phi \leqslant 1/2$. Let V be an $N \times N$ GOE matrix whose eigenvalues we denote by $\lambda_1^V \leqslant \cdots \leqslant \lambda_N^V$. Then there is a $\delta > 0$ such that for any s we have*

$$\mathbb{P}^V \Big(N^{2/3}(\lambda_N^V - 2) \leqslant s - N^{-\delta} \Big) - N^{-\delta} \; \leqslant \; \mathbb{P}^A \big(N^{2/3}(\mu_{N-1} - 2) \leqslant s \big) \; \leqslant \; \mathbb{P}^V \Big(N^{2/3}(\lambda_N^V - 2) \leqslant s + N^{-\delta} \Big) + N^{-\delta} \tag{2.11}$$

as well as

$$\mathbb{P}^V \Big(N^{2/3}(\lambda_1^V + 2) \leqslant s - N^{-\delta} \Big) - N^{-\delta} \; \leqslant \; \mathbb{P}^A \big(N^{2/3}(\mu_1 + 2) \leqslant s \big) \; \leqslant \; \mathbb{P}^V \Big(N^{2/3}(\lambda_1^V + 2) \leqslant s + N^{-\delta} \Big) + N^{-\delta} \,, \tag{2.12}$$

for $N \geqslant N_0$, where N_0 is independent of s. Here $\mathbb{P}^V$ denotes the law of the GOE matrix V, and $\mathbb{P}^A$ the law of the sparse matrix A.

Remark 2.8. Theorem 6.4 can be easily extended to correlation functions of a finite collection of extreme eigenvalues.

Remark 2.9. The GOE distribution function $F_1(s) := \lim_N \mathbb{P}^V \big(N^{2/3}(\lambda_N^V - 2) \leqslant s \big)$ of the largest eigenvalue of V has been identified by Tracy and Widom [36, 37], and can be computed in terms of Painlevé equations. A similar result holds for the smallest eigenvalue λ_1^V of V.

Remark 2.10. A result analogous to Theorem 2.7 holds for the extreme eigenvalues of the centred sparse matrix H; see (6.15) below.

We conclude this section by giving a sketch of the origin of the restriction $\phi > 1/3$ in Theorem 2.5. To simplify the outline of the argument, we set $\beta = 0$ in Theorem 2.5 and ignore any powers of N^{ε}. The proof of Theorem 2.5 is based on an analysis of the local relaxation properties of the marginal Dyson Brownian motion, obtained from the usual Dyson Brownian motion by integrating out the largest eigenvalue μ_N. As an input, we need the bound

$$Q \;:=\; \mathbb{E}\sum_{\alpha=1}^{N-1}|\mu_\alpha-\gamma_\alpha|^2 \;\leqslant\; N^{1-4\phi}, \tag{2.13}$$

where γ_α denotes the classical location of the α-th eigenvalue (see (3.15) below). The bound (2.13) was proved in [11]. In that paper we prove, roughly, that $|\mu_\alpha-\gamma_\alpha| \leqslant q^{-2} \leqslant N^{-2\phi}$, from which (2.13) follows. The precise form is given in (3.16). We then take an arbitrary initial sparse matrix ensemble A_0 and evolve it according to the Dyson Brownian motion up to a time $\tau = N^{-\rho}$, for some $\rho > 0$. We prove that the local spectral statistics, in the first $N-1$ eigenvalues, of the evolved ensemble A_τ at time τ coincide with those of a GOE matrix V, provided that

$$Q\tau^{-1} \;=\; QN^{\rho} \;\ll\; 1\,. \tag{2.14}$$

The precise statement is given in (4.9). This gives us the condition

$$1-4\phi+\rho<0\,. \tag{2.15}$$

Next, we compare the local spectral statistics of a given Erdős-Rényi matrix A with those of the time-evolved ensemble A_τ by constructing an appropriate initial A_0, chosen so that the first four moments of A and A_τ are close. More precisely, by comparing Green functions, we prove that the local spectral statistics of A and A_τ coincide if the first three moments of the entries of A and A_τ coincide and their fourth moments differ by at most $N^{-2-\delta}$ for some $\delta>0$. (See Proposition 5.2.) Given A we find, by explicit construction, a sparse matrix A_0 such that the first three moments of the entries of A_τ are equal to those of A, and their fourth moments differ by at most $N^{-1-2\phi}\tau = N^{-1-2\phi-\rho}$; see (5.6). Thus the local spectral statistics of A and A_τ coincide provided that

$$1-2\phi-\rho<0\,. \tag{2.16}$$

From the two conditions (2.15) and (2.16) we find that the local spectral statistics of A and V coincide provided that $\phi > 1/3$.

3. The strong local semicircle law and eigenvalue locations

In this preliminary section we collect the main notations and tools from the companion paper [11] that we shall need for the proofs. Throughout this paper we shall make use of the parameter

$$\xi \;\equiv\; \xi_N \;:=\; 5\log\log N\,, \tag{3.1}$$

which will keep track of powers of $\log N$ and probabilities of high-probability events. Note that in [11], ξ was a free parameter. In this paper we choose the special form (3.1) for simplicity.

We introduce the spectral parameter

$$z \;=\; E+\mathrm{i}\eta$$

where $E\in\mathbb{R}$ and $\eta>0$. Let $\Sigma\geqslant 3$ be a fixed but arbitrary constant and define the domain

$$D_L \;:=\; \big\{z\in\mathbb{C} : |E|\leqslant\Sigma\,,\,(\log N)^L N^{-1}\leqslant\eta\leqslant 3\big\}\,, \tag{3.2}$$

with a parameter $L \equiv L(N)$ that always satisfies

$$L \;\geqslant\; 8\xi\,. \tag{3.3}$$

For $\operatorname{Im} z > 0$ we define the Stieltjes transform of the local semicircle law

$$m_{sc}(z) \;:=\; \int_{\mathbb{R}} \frac{\varrho_{sc}(x)}{x - z}\,\mathrm{d}x\,, \tag{3.4}$$

where the density ϱ_{sc} was defined in (2.10). The Stieltjes transform $m_{sc}(z) \equiv m_{sc}$ may also be characterized as the unique solution of

$$m_{sc} + \frac{1}{z + m_{sc}} \;=\; 0 \tag{3.5}$$

satisfying $\operatorname{Im} m_{sc}(z) > 0$ for $\operatorname{Im} z > 0$. This implies that

$$m_{sc}(z) \;=\; \frac{-z + \sqrt{z^2 - 4}}{2}\,, \tag{3.6}$$

where the square root is chosen so that $m_{sc}(z) \sim -z^{-1}$ as $z \to \infty$. We define the resolvent of A through

$$G(z) \;:=\; (A - z)^{-1}\,,$$

as well as the Stieltjes transform of the empirical eigenvalue density

$$m(z) \;:=\; \frac{1}{N}\operatorname{Tr} G(z)\,.$$

For $x \in \mathbb{R}$ we define the distance κ_x to the spectral edge through

$$\kappa_x \;:=\; \big||x| - 2\big|\,. \tag{3.7}$$

At this point we warn the reader that we depart from our conventions in [11]. In that paper, the quantities $G(z)$ and $m(z)$ defined above in terms of A bore a tilde to distinguish them from the same quantities defined in terms of H. In this paper we drop the tilde, as we shall not need resolvents defined in terms of H.

We shall frequently have to deal with events of very high probability, for which the following definition is useful. It is characterized by two positive parameters, ξ and ν, where ξ is given by (3.1).

Definition 3.1 (High probability events). *We say that an N-dependent event Ω holds with* (ξ, ν)-high probability *if*

$$\mathbb{P}(\Omega^c) \;\leqslant\; \mathrm{e}^{-\nu(\log N)^{\xi}} \tag{3.8}$$

for $N \geqslant N_0(\nu)$.

Similarly, for a given event Ω_0, we say that Ω holds with (ξ, ν)-high probability on Ω_0 *if*

$$\mathbb{P}(\Omega_0 \cap \Omega^c) \;\leqslant\; \mathrm{e}^{-\nu(\log N)^{\xi}}$$

for $N \geqslant N_0(\nu)$.

Remark 3.2. In the following we shall not keep track of the explicit value of ν; in fact we allow ν to decrease from one line to another without introducing a new notation. All of our results will hold for $\nu \leqslant \nu_0$, where ν_0 depends only on the constants C in Definition 2.1 and the parameter Σ in (3.2).

Theorem 3.3 (Local semicircle law [11]). *Suppose that A satisfies Definition 2.2 with the condition (2.7) replaced with*

$$0 \;\leqslant\; f \;\leqslant\; N^C\,. \tag{3.9}$$

Moreover, assume that

$$q \;\geqslant\; (\log N)^{120\xi}\,, \tag{3.10}$$

$$L \;\geqslant\; 120\xi\,. \tag{3.11}$$

Then there is a constant $\nu > 0$, depending on Σ and the constants C in (2.4) and (2.5), such that the following holds.

We have the local semicircle law: the event

$$\bigcap_{z \in D_L} \left\{ \big|m(z) - m_{sc}(z)\big| \leqslant (\log N)^{40\xi} \left(\min\left\{ \frac{(\log N)^{40\xi}}{\sqrt{\kappa_E + \eta}} \frac{1}{q^2}, \frac{1}{q} \right\} + \frac{1}{N\eta} \right) \right\} \tag{3.12}$$

holds with (ξ,ν)-high probability. Moreover, we have the following estimate on the individual matrix elements of G. If instead of (3.9) f satisfies

$$0 \;\leqslant\; f \;\leqslant\; C_0 N^{1/2}\,, \tag{3.13}$$

for some constant C_0, then the event

$$\bigcap_{z \in D_L} \left\{ \max_{1 \leqslant i,j \leqslant N} \big|G_{ij}(z) - \delta_{ij} m_{sc}(z)\big| \leqslant (\log N)^{40\xi} \left(\frac{1}{q} + \sqrt{\frac{\operatorname{Im} m_{sc}(z)}{N\eta}} + \frac{1}{N\eta} \right) \right\} \tag{3.14}$$

holds with (ξ,ν)-high probability.

Next, we recall that the $N-1$ first eigenvalues of A are close the their classical locations predicted by the semicircle law. Let $n_{sc}(E) := \int_{-\infty}^E \varrho_{sc}(x)\,\mathrm{d}x$ denote the integrated density of the local semicircle law. Denote by γ_α the classical location of the α-th eigenvalue, defined through

$$n_{sc}(\gamma_\alpha) \;=\; \frac{\alpha}{N} \qquad \text{for} \quad \alpha = 1, \dots, N\,. \tag{3.15}$$

The following theorem compares the locations of the eigenvalues $\mu_1, \dots, \mu_{N-1}$ to their classical locations $\gamma_1, \dots, \gamma_{N-1}$.

Theorem 3.4 (Eigenvalue locations [11]). *Suppose that A satisfies Definition 2.2, and let ϕ be an exponent satisfying $0 < \phi \leqslant 1/2$, and set $q = N^\phi$. Then there is a constant $\nu > 0$ – depending on Σ and the constants C in (2.4), (2.5), and (2.7) – as well as a constant $C > 0$ such that the following holds.*

We have with (ξ,ν)-high probability that

$$\sum_{\alpha=1}^{N-1} |\mu_\alpha - \gamma_\alpha|^2 \;\leqslant\; (\log N)^{C\xi} \left(N^{1-4\phi} + N^{4/3-8\phi} \right). \tag{3.16}$$

Moreover, for all $\alpha = 1, \dots, N-1$ we have with (ξ,ν)-high probability that

$$|\mu_\alpha - \gamma_\alpha| \;\leqslant\; (\log N)^{C\xi} \left(N^{-2/3} \Big[\widehat{\alpha}^{-1/3} + \mathbf{1}\Big(\widehat{\alpha} \leqslant (\log N)^{C\xi}(1 + N^{1-3\phi}) \Big) \Big] + N^{2/3-4\phi} \widehat{\alpha}^{-2/3} + N^{-2\phi} \right), \tag{3.17}$$

where we abbreviated $\widehat{\alpha} := \min\{\alpha, N-\alpha\}$.

Remark 3.5. Under the assumption $\phi \geqslant 1/3$ the estimate (3.17) simplifies to

$$|\mu_\alpha - \gamma_\alpha| \;\leqslant\; (\log N)^{C\xi}\Big(N^{-2/3}\widehat{\alpha}^{-1/3} + N^{-2\phi}\Big)\,, \tag{3.18}$$

which holds with (ξ,ν)-high probability.

Finally, we record two basic results from [11] for later reference. From [11], Lemmas 4.4 and 6.1, we get, with (ξ,ν)-high probability,

$$\max_{1\leqslant\alpha\leqslant N}|\lambda_\alpha| \;\leqslant\; 2 + (\log N)^{C\xi}\Big(q^{-2} + N^{-2/3}\Big)\,, \qquad \max_{1\leqslant\alpha\leqslant N-1}|\mu_\alpha| \;\leqslant\; 2 + (\log N)^{C\xi}\Big(q^{-2} + N^{-2/3}\Big)\,. \tag{3.19}$$

Moreover, from [11], Theorem 6.2, we get, with (ξ,ν)-high probability,

$$\mu_N \;=\; f + \frac{1}{f} + o(1)\,. \tag{3.20}$$

In particular, using (2.7) we get, with (ξ,ν)-high probability,

$$2 + \sigma \;\leqslant\; \mu_N \;\leqslant\; N^C\,, \tag{3.21}$$

where $\sigma > 0$ is a constant spectral gap depending only on the constant ε_0 from (2.7).

4. Local ergodicity of the marginal Dyson Brownian motion

In Sections 4 and 5 we give the proof of Theorem 2.5. Throughout Sections 4 and 5 it is convenient to adopt a slightly different notation for the eigenvalues of A. In these two sections we shall consistently use $x_1 \leqslant \cdots \leqslant x_N$ to denote the ordered eigenvalues of A, instead of $\mu_1 \leqslant \cdots \leqslant \mu_N$ used in the rest of this paper. We abbreviate the collection of eigenvalues by $\mathbf{x} = (x_1,\dots,x_N)$.

The main tool in the proof of Theorem 2.5 is the marginal Dyson Brownian motion, obtained from the usual Dyson Brownian motion of the eigenvalues $\mathbf{x}$ by integrating out the largest eigenvalue x_N. In this section we establish the local ergodicity of the marginal Dyson Brownian and derive an upper bound on its local relaxation time.

Let $A_0 = (a_{ij,0})_{ij}$ be a matrix satisfying Definition 2.2 with constants $q_0 \geqslant N^\phi$ and $f_0 \geqslant 1+\varepsilon_0$. Let $(B_{ij,t})_{ij}$ be a symmetric matrix of independent Brownian motions, whose off-diagonal entries have variance t and diagonal entries variance $2t$. Let the matrix $A_t = (a_{ij,t})_{ij}$ satisfy the stochastic differential equation

$$\mathrm{d}a_{ij} \;=\; \frac{\mathrm{d}B_{ij}}{\sqrt{N}} - \frac{1}{2}a_{ij}\,\mathrm{d}t\,. \tag{4.1}$$

It is easy to check that the distribution of A_t is equal to the distribution of

$$\mathrm{e}^{-t/2}A_0 + (1-\mathrm{e}^{-t})^{1/2}V\,, \tag{4.2}$$

where V is a GOE matrix independent of A_0.

Let ρ be a constant satisfying $0 < \rho < 1$ to be chosen later. In the following we shall consider times t in the interval $[t_0,\tau]$, where

$$t_0 \;:=\; N^{-\rho-1}\,, \qquad \tau \;:=\; N^{-\rho}\,.$$

One readily checks that, for any fixed ρ as above, the matrix A_t satisfies Definition 2.2, with constants

$$f_t \;=\; f(1+O(N^{-\delta_0})) \;\geqslant\; 1+\frac{\varepsilon_0}{2}\,, \qquad q_t \;\sim\; q_0 \;\geqslant\; N^{\phi}\,,$$

where all estimates are uniform for $t \in [t_0, \tau]$. Denoting by $x_{N,t}$ the largest eigenvalue of A_t, we get in particular from (3.21) that

$$\mathbb{P}\Big(\exists\, t \in [t_0,\tau] : x_{N,t} \notin [2+\sigma, N^C]\Big) \;\leqslant\; \mathrm{e}^{-\nu(\log N)^{\xi}} \tag{4.3}$$

for some $\sigma > 0$ and $C > 0$.

From now on we shall never use the symbols f_t and q_t in their above sense. The only information we shall need about x_N is (4.3). In this section we shall not use any information about q_t, and in Section 5 we shall only need that $q_t \geqslant cN^{\phi}$ uniformly in t. Throughout this section f_t will denote the joint eigenvalue density evolved under the Dyson Brownian motion. (See Definition 4.1 below.)

It is well known that the eigenvalues $\mathbf{x}_t$ of A_t satisfy the stochastic differential equation (Dyson Brownian motion)

$$\mathrm{d}x_i \;=\; \frac{\mathrm{d}B_i}{\sqrt{N}} + \bigg(-\frac{1}{4}x_i + \frac{1}{2N}\sum_{j\neq i}\frac{1}{x_i - x_j}\bigg)\mathrm{d}t \qquad \text{for} \quad i = 1,\dots,N\,, \tag{4.4}$$

where $B_1,\dots,B_N$ is a family of independent standard Brownian motions.

In order to describe the law of V, we define the equilibrium Hamiltonian

$$\mathcal{H}(\mathbf{x}) \;:=\; \sum_i \frac{1}{4}x_i^2 - \frac{1}{N}\sum_{i<j}\log|x_i - x_j| \tag{4.5}$$

and denote the associated probability measure by

$$\mu^{(N)}(\mathrm{d}\mathbf{x}) \;\equiv\; \mu(\mathrm{d}\mathbf{x}) \;:=\; \frac{1}{Z}\mathrm{e}^{-N\mathcal{H}(\mathbf{x})}\,\mathrm{d}\mathbf{x}\,, \tag{4.6}$$

where Z is a normalization. We shall always consider the restriction of μ to the domain

$$\Sigma_N \;:=\; \{\mathbf{x} : x_1 < \cdots < x_N\}\,,$$

i.e. a factor $\mathbf{1}(\mathbf{x} \in \Sigma_N)$ is understood in expressions like the right-hand side of (4.6); we shall usually omit it. The law of the ordered eigenvalues of the GOE matrix V is μ.

Define the Dirichlet form D_μ and the associated generator L through

$$D_\mu(f) \;=\; -\int f(Lf)\,\mathrm{d}\mu \;:=\; \frac{1}{2N}\int |\nabla f|^2\,\mathrm{d}\mu\,, \tag{4.7}$$

where f is a smooth function of compact support on Σ_N. One may easily check that

$$L \;=\; \sum_i \frac{1}{2N}\partial_i^2 + \sum_i\bigg(-\frac{1}{4}x_i + \frac{1}{2N}\sum_{j\neq i}\frac{1}{x_i - x_j}\bigg)\partial_i\,,$$

and that L is the generator of the Dyson Brownian motion (4.4). More precisely, the law of $\mathbf{x}_t$ is given by $f_t(\mathbf{x})\,\mu(\mathrm{d}\mathbf{x})$, where f_t solves $\partial_t f_t = Lf_t$ and $f_0(\mathbf{x})\mu(\mathrm{d}\mathbf{x})$ is the law of $\mathbf{x}_0$.

Definition 4.1. *Let f_t to denote the solution of $\partial_t f_t = L f_t$ satisfying $f_t|_{t=0} = f_0$. It is well known that this solution exists and is unique, and that Σ_N is invariant under the Dyson Brownian motion, i.e. if f_0 is supported in Σ_N, so is f_t for all $t \geqslant 0$. For a precise formulation of these statements and their proofs, see e.g. Appendices A and B in [16]. In Appendix A, we present a new, simpler and more general, proof.*

Theorem 4.2. *Fix $n \geqslant 1$ and let $\mathbf{m} = (m_1, \dots, m_n) \in \mathbb{N}^n$ be an increasing family of indices. Let $G : \mathbb{R}^n \to \mathbb{R}$ be a continuous function of compact support and set*

$$\mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) \;:=\; G\big(N(x_i - x_{i+m_1}), N(x_{i+m_1} - x_{i+m_2}), \dots, N(x_{i+m_{n-1}} - x_{i+m_n})\big)\,.$$

Let $\gamma_1, \dots, \gamma_{N-1}$ denote the classical locations of the first $N-1$ eigenvalues, as defined in (3.15), *and set*

$$Q \;:=\; \sup_{t \in [t_0, \tau]} \sum_{i=1}^{N-1} \int (x_i - \gamma_i)^2 f_t \, \mathrm{d}\mu\,. \tag{4.8}$$

Choose an $\varepsilon > 0$. Then for any ρ satisfying $0 < \rho < 1$ there exists a $\bar\tau \in [\tau/2, \tau]$ such that, for any $J \subset \{1, 2, \dots, N - m_n - 1\}$, we have

$$\left| \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, f_{\bar\tau} \, \mathrm{d}\mu - \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, \mathrm{d}\mu^{(N-1)} \right| \;\leqslant\; C N^{\varepsilon} \sqrt{\frac{N^{1+\rho} Q + N^{\rho}}{|J|}} \tag{4.9}$$

for all $N \geqslant N_0(\rho)$. Here $\mu^{(N-1)}$ is the equilibrium measure of $(N-1)$ eigenvalues (GOE).

Note that, by definition, the observables $\mathcal{G}_{i,\mathbf{m}}$ in (4.9) only depend on the eigenvalues $x_1, \dots, x_{N-1}$.

The rest of this section is devoted to the proof of Theorem 4.2. We begin by introducing a pseudo equilibrium measure. Abbreviate

$$R \;:=\; \sqrt{\tau N^{-\varepsilon}} \;=\; N^{-\rho/2 - \varepsilon/2}$$

and define

$$W(\mathbf{x}) \;:=\; \sum_{i=1}^{N} \frac{1}{2R^2} (x_i - \gamma_i)^2\,.$$

Here we set $\gamma_N := 2 + \sigma$ for convenience, but one may easily check that the proof remains valid for any larger choice of γ_N. Define the probability measure

$$\omega(\mathrm{d}\mathbf{x}) \;:=\; \psi(\mathbf{x}) \, \mu(\mathrm{d}\mathbf{x}) \qquad \text{where} \quad \psi(\mathbf{x}) \;:=\; \frac{Z}{\widetilde{Z}} \mathrm{e}^{-N W(\mathbf{x})}\,.$$

Next, we consider marginal quantities obtained by integrating out the largest eigenvalue x_N. To that end we write

$$\mathbf{x} \;=\; (\widehat{x}, x_N)\,, \qquad \widehat{x} \;=\; (x_1, \dots, x_{N-1})$$

and denote by $\widehat{\omega}(\mathrm{d}\widehat{x})$ the marginal measure of ω obtained by integrating out x_N. By a slight abuse of notation, we sometimes make use of functions μ, ω, and $\widehat{\omega}$, defined as the densities (with respect to Lebesgue measure) of their respective measures. Thus,

$$\mu(\mathbf{x}) \;=\; \frac{1}{Z} \mathrm{e}^{-N\mathcal{H}(\mathbf{x})}\,, \qquad \omega(\mathbf{x}) \;=\; \frac{1}{\widetilde{Z}} \mathrm{e}^{-N\mathcal{H}(\mathbf{x}) - N W(\mathbf{x})}\,, \qquad \widehat{\omega}(\widehat{x}) \;=\; \int_{x_{N-1}}^{\infty} \omega(\widehat{x}, x_N) \, \mathrm{d}x_N\,.$$

For any function $h(\mathbf{x})$ we introduce the conditional expectation

$$\langle h \rangle(\widehat{x}) \;:=\; \mathbb{E}^{\omega}[h|\widehat{x}] \;=\; \frac{\int_{x_{N-1}}^{\infty} h(\widehat{x}, x_N)\,\omega(\widehat{x}, x_N)\,\mathrm{d}x_N}{\widehat{\omega}(\widehat{x})}\,.$$

Throughout the following, we write $g_t := f_t/\psi$. In order to avoid pathological behaviour of the extreme eigenvalues, we introduce cutoffs. Let σ be the spectral gap from (4.3), and choose $\theta_1, \theta_2, \theta_3 \in [0,1]$ to be smooth functions that satisfy

$$\theta_1(x_1) \;=\; \begin{cases} 0 & \text{if } x_1 \leqslant -4 \\ 1 & \text{if } x_1 \geqslant -3 \end{cases},$$

$$\theta_2(x_{N-1}) \;=\; \begin{cases} 1 & \text{if } x_{N-1} \leqslant 2 + \frac{\sigma}{5} \\ 0 & \text{if } x_{N-1} \geqslant 2 + \frac{2\sigma}{5} \end{cases},$$

$$\theta_3(x_N) \;=\; \begin{cases} 0 & \text{if } x_N \leqslant 2 + \frac{3\sigma}{5} \\ 1 & \text{if } x_N \geqslant 2 + \frac{4\sigma}{5} \end{cases}.$$

Define $\theta \equiv \theta(x_1, x_{N-1}, x_N) = \theta_1(x_1)\,\theta_2(x_{N-1})\,\theta_3(x_N)$. One easily finds that

$$\frac{|\nabla\theta|^2}{\theta} \;\leqslant\; C\mathbf{1}(-4 \leqslant x_1 \leqslant -3) + C\mathbf{1}\biggl(\frac{\sigma}{2} \leqslant x_{N-1} - 2 \leqslant \frac{2\sigma}{5}\biggr) + C\mathbf{1}\biggl(\frac{3\sigma}{5} \leqslant x_N - 2 \leqslant \frac{4\sigma}{5}\biggr), \tag{4.10}$$

where the left-hand side is understood to vanish outside the support of θ.

Define the density

$$h_t \;:=\; \frac{1}{\widehat{Z}_t}\,\theta g_t\,, \qquad \widehat{Z}_t \;:=\; \int \theta g_t\,\mathrm{d}\omega\,.$$

If ν is a probability measure and q a density such that $q\nu$ is also a probability measure, we define the entropy

$$S_\nu(q) \;:=\; \int q \log q\,\mathrm{d}\nu\,.$$

The following result is our main tool for controlling the local ergodicity of the marginal Dyson Brownian motion.

PROPOSITION 4.3. *Suppose that*

$$\textit{(i)}\quad S_\mu(f_{t_0}) \;\leqslant\; N^C\,, \tag{4.11}$$

$$\textit{(ii)}\quad \sup_{t \in [t_0,\tau]} \int \biggl[\mathbf{1}(x_1 \leqslant -3) + \mathbf{1}\biggl(x_{N-1} \geqslant 2 + \frac{\sigma}{5}\biggr) + \mathbf{1}\biggl(x_N \leqslant 2 + \frac{4\sigma}{5}\biggr)\biggr] f_t\,\mathrm{d}\mu \;\leqslant\; \mathrm{e}^{-\nu(\log N)^\xi}\,, \tag{4.12}$$

$$\textit{(iii)}\quad \sup_{t \in [t_0,\tau]}\, \sup_{\widehat{x} \in \Sigma_{N-1}} (\theta_1\theta_2)(\widehat{x})\bigl|\log\langle\theta g_t\rangle(\widehat{x})\bigr|^2 \;\leqslant\; N^C\,. \tag{4.13}$$

Then for $t \in [t_0, \tau]$ we have

$$\partial_t S_{\widehat{\omega}}(\langle h\rangle) \;\leqslant\; -D_{\widehat{\omega}}\bigl(\sqrt{\langle h\rangle}\bigr) + S_{\widehat{\omega}}(\langle h\rangle)\,\mathrm{e}^{-c(\log N)^\xi} + CNQR^{-4} + C\,. \tag{4.14}$$

Proof. First we note that

$$\widehat{Z}_t \;=\; \int \theta f_t \,\mathrm{d}\mu \;=\; 1 - O\big(\mathrm{e}^{-\nu(\log N)^\xi}\big) \tag{4.15}$$

uniformly for $t \in [t_0, \tau]$, by (4.12). Dropping the time index to avoid cluttering the notation, we find

$$\partial_t S_{\widehat{\omega}}(\langle h \rangle) \;=\; \partial_t \int \frac{\langle \theta g \rangle}{\widehat{Z}} \log\langle \theta g \rangle \,\mathrm{d}\widehat{\omega} - \partial_t \log \widehat{Z} \;=\; \frac{1}{\widehat{Z}} \partial_t \int \theta g \,\log\langle \theta g \rangle \,\mathrm{d}\omega - \big(1 + \log \widehat{Z} + S_{\widehat{\omega}}(\langle h \rangle)\big)\, \partial_t \log \widehat{Z}\,.$$

We find that

$$\partial_t \widehat{Z} \;=\; \int \theta (Lf) \,\mathrm{d}\mu \;=\; -\frac{1}{2N} \int \nabla\theta \cdot \nabla f \,\mathrm{d}\mu \;\leqslant\; \bigg(\frac{1}{N} \int |\nabla \theta|^2 \, f \,\mathrm{d}\mu\bigg)^{1/2} D_\mu(\sqrt{f})^{1/2}\,.$$

Bounding the Dirichlet form in terms of the entropy (see e.g. [10], Theorem 3.2), we find that

$$D_\mu(\sqrt{f_t}) \;\leqslant\; \frac{2}{t} S_\mu(f_{t_0}) \;\leqslant\; N^C\,, \tag{4.16}$$

by (4.11). Using (4.10) we therefore find

$$\partial_t \widehat{Z} \;\leqslant\; N^C \mathrm{e}^{-c(\log N)^\xi}\,. \tag{4.17}$$

Thus we have

$$\partial_t S_{\widehat{\omega}}(\langle h \rangle) \;\leqslant\; 2 \partial_t \int \theta g \,\log\langle \theta g \rangle \,\mathrm{d}\omega + \big(1 + S_{\widehat{\omega}}(\langle h \rangle)\big) N^C \mathrm{e}^{-c(\log N)^\xi}\,. \tag{4.18}$$

We therefore need to estimate

$$\partial_t \int \theta g \,\log\langle \theta g \rangle \,\mathrm{d}\omega \;=\; \int \theta (Lf) \log\langle \theta g \rangle \,\mathrm{d}\mu + \int \langle \theta g \rangle \,\frac{\partial_t \langle \theta g \rangle}{\langle \theta g \rangle} \,\mathrm{d}\widehat{\omega}\,. \tag{4.19}$$

The second term of (4.19) is given by

$$\int \partial_t \langle \theta g \rangle \,\mathrm{d}\widehat{\omega} \;=\; \int \theta (Lf) \,\mathrm{d}\mu \;=\; \partial_t \widehat{Z}\,.$$

Therefore (4.18) yields

$$\partial_t S_{\widehat{\omega}}(\langle h \rangle) \;\leqslant\; 2 \int \theta (Lf) \log\langle \theta g \rangle \,\mathrm{d}\mu + (1 + S_{\widehat{\omega}}(\langle h \rangle)) N^C \mathrm{e}^{-c(\log N)^\xi}\,. \tag{4.20}$$

The first term of (4.20) is given by

$$-\frac{1}{N} \int \nabla f \cdot \nabla\big(\theta \log\langle \theta g \rangle\big) \,\mathrm{d}\mu \;=\; -\frac{1}{N} \int \nabla(\theta f) \cdot \nabla\big(\log\langle \theta g \rangle\big) \,\mathrm{d}\mu + \mathcal{E}_1 + \mathcal{E}_2\,, \tag{4.21}$$

where we defined

$$\mathcal{E}_1 \;:=\; \frac{1}{N} \int \nabla \theta \cdot \nabla(\log\langle \theta g \rangle)\, f \,\mathrm{d}\mu\,, \qquad \mathcal{E}_2 \;:=\; -\frac{1}{N} \int \nabla\theta \cdot \nabla f \,\log\langle \theta g \rangle \,\mathrm{d}\mu\,.$$

Next, we estimate the error terms $\mathcal{E}_1$ and $\mathcal{E}_2$. Using (4.10) we get

$$\begin{aligned}
\mathcal{E}_1 \;=\; \frac{1}{N}\int \frac{\nabla\theta}{\sqrt{\theta}}\cdot\nabla(\log\langle\theta g\rangle)\,\sqrt{\theta} f\,\mathrm{d}\mu \;&\leqslant\; \biggl(\int \frac{|\nabla\theta|^2}{\theta} f\,\mathrm{d}\mu\biggr)^{1/2}\biggl(\int \frac{|\nabla\langle\theta g\rangle|^2}{\langle\theta g\rangle^2}\,\theta f\,\mathrm{d}\mu\biggr)^{1/2}\\
&\leqslant\; \mathrm{e}^{-\nu(\log N)^\xi}\biggl(\int \frac{|\nabla\langle\theta g\rangle|^2}{\langle\theta g\rangle^2}\,\langle\theta g\rangle\,\mathrm{d}\widehat{\omega}\biggr)^{1/2} \;\leqslant\; \mathrm{e}^{-c(\log N)^\xi} + \mathrm{e}^{-c(\log N)^\xi} D_{\widehat{\omega}}\bigl(\sqrt{\langle h\rangle}\bigr)\,,
\end{aligned}$$

where we used (4.15). Similarly, we find

$$\mathcal{E}_2 \;\leqslant\; \biggl(\int |\nabla\theta|^2\,\bigl|\log\langle\theta g\rangle\bigr|^2 f\,\mathrm{d}\mu\biggr)^{1/2}\biggl(\int \frac{|\nabla f|^2}{f}\,\mathrm{d}\mu\biggr)^{1/2}.$$

Using (4.10), (4.13), and (4.16) we therefore get

$$\mathcal{E}_2 \;\leqslant\; N^C\biggl(\int \frac{|\nabla\theta|^2}{\theta}\,\theta\bigl|\log\langle\theta g\rangle\bigr|^2 f\,\mathrm{d}\mu\biggr)^{1/2} \;\leqslant\; N^C\mathrm{e}^{-c(\log N)^\xi}\,.$$

Having dealt with the error terms $\mathcal{E}_1$ and $\mathcal{E}_2$, we compute the first term on the right-hand side of (4.21),

$$-\frac{1}{N}\int\nabla(\theta f)\cdot\nabla\bigl(\log\langle\theta g\rangle\bigr)\,\mathrm{d}\mu \;=\; -\frac{1}{N}\int\nabla_{\widehat{x}}(\theta g)\cdot\nabla_{\widehat{x}}\bigl(\log\langle\theta g\rangle\bigr)\,\psi\,\mathrm{d}\mu - \frac{1}{N}\int\nabla_{\widehat{x}}(\log\psi)\cdot\nabla_{\widehat{x}}\bigl(\log\langle\theta g\rangle\bigr)\,\theta g\psi\,\mathrm{d}\mu\,. \tag{4.22}$$

The second term of (4.22) is bounded by

$$\begin{aligned}
\frac{\eta^{-1}}{N}\int|\nabla_{\widehat{x}}\log\psi|^2 f\,\mathrm{d}\mu + \frac{\eta}{N}\int\frac{|\nabla\,\langle\theta g\rangle|^2}{\langle\theta g\rangle^2}\,\langle\theta g\rangle\,\mathrm{d}\widehat{\omega} \;&\leqslant\; \eta^{-1}N\int\frac{1}{R^4}\sum_{i=1}^{N-1}(x_i-\gamma_i)^2 f\,\mathrm{d}\mu + 4\eta D_{\widehat{\omega}}(\sqrt{\langle h\rangle})\\
&\leqslant\; \frac{NQ}{\eta R^4} + 4\eta D_{\widehat{\omega}}(\sqrt{\langle h\rangle})\,,
\end{aligned}$$

where $\eta > 0$.

The first term of (4.22) is equal to

$$-\frac{1}{N}\int\langle\nabla_{\widehat{x}}(\theta g)\rangle\cdot\nabla_{\widehat{x}}\bigl(\log\langle\theta g\rangle\bigr)\,\mathrm{d}\widehat{\omega}\,.$$

A simple calculation shows that

$$\langle\nabla_{\widehat{x}}(\theta g)\rangle \;=\; \nabla_{\widehat{x}}\langle\theta g\rangle - \bigl\langle\theta g\nabla_{\widehat{x}}\log\omega\bigr\rangle + \langle\theta g\rangle\,\langle\nabla_{\widehat{x}}\log\omega\rangle\,,$$

so that the first term of (4.22) becomes

$$\begin{aligned}
-\frac{1}{N}\int\nabla_{\widehat{x}}\langle\theta g\rangle\cdot\nabla_{\widehat{x}}\bigl(\log\langle\theta g\rangle\bigr)\,\mathrm{d}\widehat{\omega} &+ \frac{1}{N}\int\Bigl(\bigl\langle\theta g\nabla_{\widehat{x}}\log\omega\bigr\rangle - \langle\theta g\rangle\,\langle\nabla_{\widehat{x}}\log\omega\rangle\Bigr)\cdot\nabla_{\widehat{x}}\bigl(\log\langle\theta g\rangle\bigr)\,\mathrm{d}\widehat{\omega}\\
&\leqslant\; -4(1-\eta)D_{\widehat{\omega}}\bigl(\sqrt{\langle h\rangle}\bigr) + \frac{1}{N\eta}\int\frac{\bigl|\bigl\langle\theta g\nabla_{\widehat{x}}\log\omega\bigr\rangle - \langle\theta g\rangle\,\langle\nabla_{\widehat{x}}\log\omega\rangle\bigr|^2}{\langle\theta g\rangle}\,\mathrm{d}\widehat{\omega}\,.
\end{aligned}$$

Using the Cauchy-Schwarz inequality $\langle ab\rangle^2 \leqslant \langle a^2\rangle \langle b^2\rangle$ we find that the second term is bounded by

$$\begin{aligned}
\frac{1}{N\eta}\int \frac{\big|\big\langle \theta g\big(\nabla_{\widehat{x}}\log\omega - \langle\nabla_{\widehat{x}}\log\omega\rangle\big)\big\rangle\big|^2}{\langle\theta g\rangle}\,\mathrm{d}\widehat{\omega} \;&\leqslant\; \frac{1}{N\eta}\int \Big\langle \theta g\big|\nabla_{\widehat{x}}\log\omega - \langle\nabla_{\widehat{x}}\log\omega\rangle\big|^2\Big\rangle\,\mathrm{d}\widehat{\omega} \\
&=\; \frac{1}{N\eta}\int \big|\nabla_{\widehat{x}}\log\omega - \langle\nabla_{\widehat{x}}\log\omega\rangle\big|^2\,\theta f\,\mathrm{d}\mu \\
&=\; \frac{1}{N\eta}\int \sum_{i=1}^{N-1}\bigg(\frac{1}{x_N - x_i} - \bigg\langle\frac{1}{x_N - x_i}\bigg\rangle\bigg)^2\,\theta f\,\mathrm{d}\mu\,.
\end{aligned}$$

Thus, we have to estimate

$$\mathcal{E}_3 \;:=\; \frac{1}{N\eta}\int\sum_{i=1}^{N-1}\bigg(\frac{1}{x_N-x_i}\bigg)^2\theta f\,\mathrm{d}\mu\,, \qquad \mathcal{E}_4 \;:=\; \frac{1}{N\eta}\int\sum_{i=1}^{N-1}\bigg\langle\frac{1}{x_N-x_i}\bigg\rangle^2\theta f\,\mathrm{d}\mu\,.$$

Since $x_N - x_i \geqslant \sigma/5$ on the support of $\theta f\mu$, one easily gets from (3.19) that

$$\mathcal{E}_3 \;\leqslant\; \frac{C}{\eta}\,.$$

In order to estimate $\mathcal{E}_4$, we write

$$\bigg\langle\frac{1}{x_N-x_i}\bigg\rangle \;=\; \bigg(\frac{\int \mathrm{d}x_N\,(x_N-x_i)\,w_i(x_N)}{\int\mathrm{d}x_N\,w_i(x_N)}\bigg)^{-1}\,,$$

where

$$w_i(x_N) \;:=\; \mathbf{1}(x_N \geqslant x_{N-1})\,\mathrm{e}^{-\frac{N}{4}x_N^2 - \frac{N}{2R^2}(x_N-\gamma_N)^2}\prod_{j\neq i,N}(x_N-x_j)\,.$$

We now claim that on the support of θ, in particular for $-4 \leqslant x_1 < x_{N-1} \leqslant 2 + 2\sigma/5$, we have

$$\frac{\int \mathrm{d}x_N\,(x_N-x_i)\,w_i(x_N)}{\int\mathrm{d}x_N\,w_i(x_N)} \;\geqslant\; c\,\gamma_N\,, \tag{4.23}$$

uniformly for $\widehat{x}\in\Sigma_{N-1}$. Indeed, writing $\widetilde{\gamma}_N := \gamma_N(1+R^{-2})$, we have on the support of θ

$$\frac{\int \mathrm{d}x_N\,(x_N-x_i)\,w_i(x_N)}{\int\mathrm{d}x_N\,w_i(x_N)} \;\geqslant\; \widetilde{\gamma}_N/2 + \frac{\int \mathrm{d}x_N\,(x_N-\widetilde{\gamma}_N)\,w_i(x_N)}{\int\mathrm{d}x_N\,w_i(x_N)}\,.$$

Moreover, the second term is nonnegative:

$$\begin{aligned}
\int \mathrm{d}x_N\,(x_N-\widetilde{\gamma}_N)\,w_i(x_N) \;&=\; -C_N(\widehat{x})\int_{x_{N-1}}^{\infty}\mathrm{d}x_N\,\bigg(\frac{\partial}{\partial x_N}\mathrm{e}^{-\frac{N}{R^2}(x_N-\widetilde{\gamma}_N)^2}\bigg)\prod_{j\neq i,N}(x_N-x_j) \\
&=\; C_N(\widehat{x})\,\mathrm{e}^{-\frac{N}{R^2}(x_{N-1}-\widetilde{\gamma}_N)^2}\prod_{j\neq i,N}(x_{N-1}-x_j) \\
&\qquad + C_N(\widehat{x})\int_{x_{N-1}}^{\infty}\mathrm{d}x_N\,\mathrm{e}^{-\frac{N}{R^2}(x_N-\widetilde{\gamma}_N)^2}\sum_{k\neq i,j}\;\prod_{j\neq i,k,N}(x_N-x_j) \\
&\geqslant\; 0\,,
\end{aligned}$$

where $C_N(\widehat{x})$ is nonnegative. This proves (4.23). Using (4.23) we get

$$\mathcal{E}_4 \;\leqslant\; \frac{C}{\eta} \int \gamma_N^{-2}\, f \,\mathrm{d}\mu \;=\; \frac{C}{\eta}\,.$$

Summarizing, we have proved that

$$\partial_t S_{\widehat{\omega}}(\langle h \rangle) \;\leqslant\; -\big(4 - 8\eta - \mathrm{e}^{-c(\log N)^\xi}\big) D_{\widehat{\omega}}\big(\sqrt{\langle h \rangle}\big) + \big(1 + S_{\widehat{\omega}}(\langle h \rangle)\big)\mathrm{e}^{-c(\log N)^\xi} + \frac{NQ}{\eta R^4} + \frac{C}{\eta}\,.$$

Choosing η small enough completes the proof. □

Next, we derive a logarithmic convexity bound for the marginal measure $\widehat{\omega}$.

Lemma 4.4. *We have that*

$$\widehat{\omega}(\widehat{x}) \;=\; \frac{1}{\widetilde{Z}}\, \mathrm{e}^{-N \widehat{\mathcal{H}}(\widehat{x})}\,,$$

where

$$\widehat{\mathcal{H}}(\widehat{x}) \;=\; -\frac{1}{N} \sum_{i<j<N} \log|x_i - x_j| + V(\widehat{x})\,, \tag{4.24}$$

and $\nabla^2 V(\widehat{x}) \geqslant R^{-2}$.

Proof. Write $\mathcal{H}(\widehat{x}, x_N) = \mathcal{H}'(\widehat{x}) + \mathcal{H}''(\widehat{x}, x_N)$ where

$$\mathcal{H}'(\widehat{x}) \;:=\; -\frac{1}{N} \sum_{i<j<N} \log|x_i - x_j|\,, \qquad \mathcal{H}''(\widehat{x}, x_N) \;:=\; -\frac{1}{N} \sum_{i<N} \log|x_N - x_i| + \sum_i \frac{1}{2R^2}(x_i - \gamma_i)^2\,.$$

By definition, we have

$$\widehat{\omega}(\widehat{x}) \;=\; \frac{1}{\widetilde{Z}} \mathrm{e}^{-N\mathcal{H}'(\widehat{x})} \int_{x_{N-1}}^{\infty} \mathrm{e}^{-N\mathcal{H}''(\widehat{x}, x_N)}\, \mathrm{d}x_N\,.$$

The main tool in our proof is the Brascamp-Lieb inequality [6]. In order to apply it, we need to extend the integration over x_N to $\mathbb{R}$ and replace the singular logarithm with a C^2-function. To that end, we introduce the approximation parameter $\delta > 0$ and define, for $\widehat{x} \in \Sigma_{N-1}$,

$$V_\delta(\widehat{x}) \;:=\; -\frac{1}{N} \log \int \exp\Bigg[\sum_{i<N} \log_\delta(x_N - x_i) - \frac{N}{2R^2} \sum_i (x_i - \gamma_i)^2\Bigg]\, \mathrm{d}x_N\,,$$

where we defined

$$\log_\delta(x) \;:=\; \mathbf{1}(x \geqslant \delta) \log x + \mathbf{1}(x < \delta)\left(\log \delta + \frac{x - \delta}{\delta} - \frac{1}{2\delta^2}(x - \delta)^2\right).$$

It is easy to check that $\log_\delta \in C^2(\mathbb{R})$, is concave, and satisfies

$$\lim_{\delta \to 0} \log_\delta(x) \;=\; \begin{cases} \log x & \text{if } x > 0 \\ -\infty & \text{if } x \leqslant 0\,. \end{cases}$$

Thus we find that $V_\delta \in C^2(\Sigma_{N-1})$ and that we have the pointwise convergence, for all $\widehat{x} \in \Sigma_{N-1}$,

$$\lim_{\delta \to 0} V_\delta(\widehat{x}) \;=\; V(\widehat{x}) \;:=\; -\frac{1}{N} \log \int_{x_{N-1}}^\infty \mathrm{e}^{-N\mathcal{H}''(\widehat{x}, x_N)} \, \mathrm{d}x_N \,,$$

where $V \in C^2(\Sigma_{N-1})$ satisfies (4.24).

Next, we claim that if $\varphi = \varphi(x,y)$ satisfies $\nabla^2 \varphi(x,y) \geqslant K$ then $\psi(x)$, defined by

$$\mathrm{e}^{-\psi(x)} \;:=\; \int \mathrm{e}^{-\varphi(x,y)} \, \mathrm{d}y \,,$$

satisfies $\nabla^2 \psi(x) \geqslant K$. In order to prove the claim, we use subscripts to denote partial derivatives and recall the Brascamp-Lieb inequality for log-concave functions (Equation 4.7 in [6])

$$\psi_{xx} \;\geqslant\; \frac{\int \big(\varphi_{xx} - \varphi_{xy} \varphi_{yy}^{-1} \varphi_{yx}\big) \, \mathrm{e}^{-\varphi} \, \mathrm{d}y}{\int \mathrm{e}^{-\varphi} \, \mathrm{d}y} \,.$$

Then the claim follows from

$$\begin{pmatrix} \varphi_{xx} & \varphi_{xy} \\ \varphi_{yz} & \varphi_{yy} \end{pmatrix}^{-1} \;\leqslant\; \frac{1}{K} \qquad \Longrightarrow \qquad \big(\varphi_{xx} - \varphi_{xy} \varphi_{yy}^{-1} \varphi_{yx}\big) \;\geqslant\; K \,.$$

Using this claim, we find that $\nabla^2 V_\delta(\widehat{x}) \geqslant R^{-2}$ for all $\widehat{x} \in \Sigma_{N-1}$. In order to prove that $\nabla^2 V(\widehat{x}) \geqslant R^{-2}$ – and hence complete the proof – it suffices to consider directional derivatives and prove the following claim. If $(\zeta_\delta)_{\delta > 0}$ is a family of functions on a neighbourhood U that converges pointwise to a C^2-function ζ as $\delta \to 0$, and if $\zeta_\delta''(x) \geqslant K$ for all $\delta > 0$ and $x \in U$, then $\zeta''(x) \geqslant K$ for all $x \in U$. Indeed, taking $\delta \to 0$ in

$$\zeta_\delta(x+h) + \zeta_\delta(x-h) - 2\zeta_\delta(x) \;=\; \int_0^h \big(\zeta_\delta''(x+\xi) + \zeta_\delta''(x-\xi)\big)(h-\xi) \, \mathrm{d}\xi \;\geqslant\; Kh^2$$

yields $\big(\zeta(x+h) + \zeta(x-h) - 2\zeta(x)\big) h^{-2} \geqslant K$, from which the claim follows by taking the limit $h \to 0$. ☐

As a first consequence of Lemma 4.4, we derive an estimate on the expectation of observables depending only on eigenvalue differences.

PROPOSITION 4.5. *Let $q \in L^\infty(\mathrm{d}\widehat{\omega})$ be probability density. Then for any $J \subset \{1, 2, \dots, N - m_n - 1\}$ and any $t > 0$ we have*

$$\left| \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, q \, \mathrm{d}\widehat{\omega} - \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, \mathrm{d}\widehat{\omega} \right| \;\leqslant\; C \sqrt{\frac{D_{\widehat{\omega}}(\sqrt{q}) \, t}{|J|}} + C \sqrt{S_{\widehat{\omega}}(q)} \, \mathrm{e}^{-ct/R^2} \,.$$

PROOF. Using Lemma 4.4, the proof of Theorem 4.3 in [16] applies with merely cosmetic changes. ☐

Another, standard, consequence of Lemma 4.4 is the logarithmic Sobolev inequality

$$S_{\widehat{\omega}}(q) \;\leqslant\; C R^2 D_{\widehat{\omega}}(\sqrt{q}) \,. \tag{4.25}$$

Using (4.25) and Proposition 4.3, we get the following estimate on the Dirichlet form.

Proposition 4.6. *Under the assumptions of Proposition 4.3, there exists a* $\bar{\tau} \in [\tau/2, \tau]$ *such that*

$$S_{\widehat{\omega}}(\langle h_{\bar{\tau}} \rangle) \leqslant CNR^{-2}Q + CR^2 , \qquad D_{\widehat{\omega}}(\sqrt{\langle h_{\bar{\tau}} \rangle}) \leqslant CNR^{-4}Q + C .$$

Proof. Combining (4.25) with (4.14) yields

$$\partial_t S_{\widehat{\omega}}(\langle h_t \rangle) \leqslant -CR^{-2} S_{\widehat{\omega}}(\langle h_t \rangle) + CNQR^{-4} + C , \tag{4.26}$$

which we integrate from t_0 to t to get

$$S_{\widehat{\omega}}(\langle h_t \rangle) \leqslant \mathrm{e}^{-CR^{-2}(t-t_0)} S_{\widehat{\omega}}(\langle h_{t_0} \rangle) + CNQR^{-2} + CR^2 . \tag{4.27}$$

Moreover, (4.15) yields

$$\begin{aligned} S_{\widehat{\omega}}(\langle h_{t_0} \rangle) \leqslant CS_{\widehat{\omega}}(\langle g_{t_0} \rangle) + \mathrm{e}^{-\nu(\log N)^\xi} &\leqslant CS_\omega(g_{t_0}) + \mathrm{e}^{-\nu(\log N)^\xi} \\ &= CS_\mu(f_{t_0}) - C \int \log \psi \, f_{t_0} \, \mathrm{d}\mu + \mathrm{e}^{-\nu(\log N)^\xi} , \end{aligned}$$

where the second inequality follows from the fact that taking marginals reduces the relative entropy; see the proof of Lemma 4.7 below for more details. Thus we get

$$S_{\widehat{\omega}}(\langle h_{t_0} \rangle) \leqslant N^C + NR^{-2}Q \leqslant N^C .$$

Thus (4.27) yields

$$S_{\widehat{\omega}}(\langle h_t \rangle) \leqslant N^C \mathrm{e}^{-CR^{-2}(t-t_0)} + CNR^{-2}Q + CR^2 \tag{4.28}$$

for $t \in [t_0, \tau]$. Integrating (4.14) from $\tau/2$ to τ therefore gives

$$\frac{2}{\tau} \int_{\tau/2}^{\tau} D_{\widehat{\omega}}(\sqrt{\langle h_t \rangle}) \, \mathrm{d}t \leqslant CNR^{-4}Q + C ,$$

and the claim follows. □

We may finally complete the proof of Theorem 4.2.

Proof of Theorem 4.2. The assumptions of Proposition 4.3 are verified in Subsection 4.1 below. Hence Propositions 4.5 and 4.6 yield

$$\left| \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, h_{\bar{\tau}} \, \mathrm{d}\omega - \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, \mathrm{d}\omega \right| \leqslant CN^\varepsilon \sqrt{\frac{N^{1+\rho}Q}{|J|}} + C \sqrt{\frac{N^{-2\phi-\rho}}{|J|}} .$$

Using (4.15) and (4.12) we get

$$\left| \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, f_{\bar{\tau}} \, \mathrm{d}\mu - \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, \mathrm{d}\omega \right| \leqslant CN^\varepsilon \sqrt{\frac{N^{1+\rho}Q}{|J|}} + C \sqrt{\frac{N^{-2\phi-\rho}}{|J|}} . \tag{4.29}$$

In order to compare the measures $\widehat{\omega}$ and $\mu^{(N-1)}$, we define the density

$$q(\mathbf{x}) \;:=\; \frac{1}{Z'}\exp\bigg\{\sum_{i<N}\frac{1}{4}x_i^2+\sum_{i<N}\frac{N}{2R^2}(x_i-\gamma_i)^2-\sum_{i<N}\log|x_N-x_i|\bigg\},$$

where Z' is a normalization chosen so that $\theta q\,\mathrm{d}\omega$ is a probability measure. It is easy to see that

$$q\,\mathrm{d}\omega \;=\; \mathrm{d}\mu^{(N-1)}\otimes\mathrm{d}g\,,$$

where $\mathrm{d}g = C\mathrm{e}^{-\frac{N}{4}x_N^2-\frac{N}{2R^2}(x_N-\gamma_N)^2}\mathrm{d}x_N$ is a Gaussian measure. Similarly to Proposition 4.5, we have

$$\bigg|\int\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}\,\theta q\,\mathrm{d}\omega-\int\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}\,\mathrm{d}\omega\bigg| \;\leqslant\; C\sqrt{\frac{D_\omega(\sqrt{\theta q})\tau}{|J|}}+C\sqrt{S_\omega(\theta q)}\,\mathrm{e}^{-c\tau/R^2}\,.$$

Thus we have to estimate

$$\begin{aligned}
D_\omega(\sqrt{\theta q}) \;&\leqslant\; \frac{C}{N}\int|\nabla\log q|^2\,\theta q\,\mathrm{d}\omega+\frac{C}{N}\int\frac{|\nabla\theta|^2}{\theta}\,q\,\mathrm{d}\omega\\
&\leqslant\; \frac{C}{N}\sum_{i<N}\int\bigg(\frac{1}{4}x_i^2+\frac{N^2}{R^4}(x_i-\gamma_i)^2+\frac{1}{(x_N-x_i)^2}\bigg)\theta q\,\mathrm{d}\omega+\frac{1}{N}\\
&\leqslant\; C+NR^{-4}\int\sum_{i<N}(x_i-\gamma_i)^2\,\mathrm{d}\mu^{(N-1)}
\end{aligned}$$

where the second inequality follows from standard large deviation results for GOE. Since $\int\sum_{i<N}(x_i-\gamma_i)^2\,\mathrm{d}\mu^{(N-1)}\leqslant CN^{-1+\varepsilon'}$ for arbitrary ε' is known to hold for GOE (see [19] where this is proved for more general Wigner matrices), we find

$$\bigg|\int\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}\,\theta q\,\mathrm{d}\omega-\int\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}\,\mathrm{d}\omega\bigg| \;\leqslant\; C\sqrt{\frac{N^{-\rho}}{|J|}}+C\sqrt{\frac{N^{\rho+2\varepsilon+\varepsilon'}}{|J|}}\,.$$

The cutoff θ can be easily removed using the standard properties of $\mathrm{d}\mu^{(N-1)}$. Choosing $\varepsilon'=\varepsilon$, replacing ε with $\varepsilon/2$, and recalling (4.29) completes the proof. □

4.1. Verifying the assumptions of Proposition 4.3. The estimate (4.11) is an immediate consequence of the following lemma.

LEMMA 4.7. *Let the entries of A_0 have the distribution ζ_0. Then for any $t>0$ we have*

$$S_\mu(f_t) \;\leqslant\; N^2\big(Nm_2(\zeta_0)-\log\big(1-\mathrm{e}^{-t}\big)\big)\,,$$

where $m_2(\zeta_0)$ is the second moment of ζ_0.

PROOF. Recall that the relative entropy is defined, for $\nu\ll\mu$, as $S(\nu|\mu) \;:=\; \int\log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,\mathrm{d}\nu$. If $\widehat{\nu}$ and $\widehat{\mu}$ are marginals of ν and μ with respect to the same variable, it is easy to check that $S(\widehat{\nu}|\widehat{\mu})\leqslant S(\nu|\mu)$. Therefore

$$S_\mu(f_t) \;=\; S(f_t\mu|\mu) \;\leqslant\; S(A_t|V) \;=\; N^2S(\zeta_t|g_{2/N})\,,$$

where ζ_t denotes the law of the off-diagonal entries of A_t, and g_λ is a standard Gaussian with variance λ (the diagonal entries are dealt with similarly). Setting $\gamma = 1 - \mathrm{e}^{-t}$, we find from (4.2) that ζ_t has probability density $\varrho_\gamma * g_{2\gamma/N}$, where ϱ_γ is the probability density of $(1-\gamma)^{1/2}\zeta_0$. Therefore Jensen's inequality yields

$$S(\zeta_t | g_{2/N}) \;=\; S\biggl(\int \mathrm{d}y\, \varrho_\gamma(y)\, g_{2\gamma/N}(\cdot - y) \,\bigg|\, g_{2/N}\biggr) \;\leqslant\; \int \mathrm{d}y\, \varrho_\gamma(y) S\bigl(g_{2\gamma/N}(\cdot - y)|g_{2/N}\bigr)\,.$$

By explicit computation one finds

$$S\bigl(g_{2\gamma/N}(\cdot - y)|g_{2/N}\bigr) \;=\; \frac{1}{2}\biggl(\frac{N}{2}y^2 - \log\gamma + \gamma - 1\biggr)\,.$$

Therefore

$$S(\zeta_t|g_{2/N}) \;\leqslant\; N m_2(\zeta_0) - \log\gamma\,,$$

and the claim follows. ☐

The estimate (4.12) follows from (4.3) and (3.19). It only remains to verify (4.13).

Lemma 4.8. *For any $t \in [t_0, \tau]$ we have*

$$(\theta_1\theta_2)(\widehat{x})\bigl|\log\langle\theta g_t\rangle(\widehat{x})\bigr|^2 \;\leqslant\; N^C\,. \tag{4.30}$$

Proof. Let ζ_t be the law of an off-diagonal entry a of A_t (the diagonal entries are treated similarly). From (4.2) we find

$$\zeta_t \;=\; \varrho_\gamma * g_{2\gamma/N}\,,$$

where $\gamma = 1 - \mathrm{e}^{-t}$, ϱ_γ is the law of $(1-\gamma)^{1/2}\zeta_0$, and g_λ is a standard Gaussian with variance λ. Using $\mathrm{d}a$ to denote Lebesgue measure, we find by explicit calculation that

$$\mathrm{e}^{-N^C - N^C a^2} \;\leqslant\; \frac{\mathrm{d}\zeta_t}{\mathrm{d}a} \;\leqslant\; \mathrm{e}^{N^C - \frac{N}{4}a^2}\,,$$

which gives

$$\mathrm{e}^{-N^C - N^C a^2} \;\leqslant\; \frac{\mathrm{d}\zeta_t}{\mathrm{d}g_{2\gamma/N}} \;\leqslant\; \mathrm{e}^{N^C}\,.$$

Therefore, the density $F_t(A)$ of the law of A with respect to the GOE measure satisfies

$$\mathrm{e}^{-N^C - N^C \operatorname{Tr} A^2} \;\leqslant\; F_t(A) \;\leqslant\; \mathrm{e}^{N^C}\,.$$

Parametrizing $A = A(\mathbf{x}, \mathbf{v})$ using the eigenvalues $\mathbf{x}$ and eigenvectors $\mathbf{v}$, the GOE measure can be written in the factorized form $\mu(\mathrm{d}\mathbf{x})P(\mathrm{d}\mathbf{v})$, where μ is defined in (4.6) and P is a probability measure. Thus we get that the density

$$f_t(\mathbf{x}) \;=\; \int F_t(\mathbf{x}, \mathbf{v})\, P(\mathrm{d}\mathbf{v})$$

satisfies

$$\mathrm{e}^{-N^C - N^C \sum_i x_i^2} \;\leqslant\; f_t(\mathbf{x}) \;\leqslant\; \mathrm{e}^{N^C}\,. \tag{4.31}$$

Next, it is easy to see that

$$\mathrm{e}^{-N^C - N^C \sum_i x_i^2} \;\leqslant\; \psi(\mathbf{x}) \;\leqslant\; \mathrm{e}^{N^C}\,. \tag{4.32}$$

Using (4.32) we may now derive an upper bound on $\langle \theta_3 g_t \rangle$:

$$\begin{aligned}
\langle \theta_3 g_t \rangle(\widehat{x}) \;&=\; \frac{\int \mathrm{d}x_N \, \theta_3(x_N) f_t(\widehat{x}, x_N) \mu(\widehat{x}, x_N)}{\int \mathrm{d}x_N \, \psi(\widehat{x}, x_N) \mu(\widehat{x}, x_N)} \\
&\leqslant\; \mathrm{e}^{N^C + N^C \sum_{i<N} x_i^2} \frac{\int \mathrm{d}x_N \, \mu(\widehat{x}, x_N)}{\int \mathrm{d}x_N \, \mathrm{e}^{-N^C x_N^2} \mu(\widehat{x}, x_N)}\,.
\end{aligned}$$

Since

$$\frac{\int \mathrm{d}x_N \, \mathrm{e}^{-N^C x_N^2} \mu(\widehat{x}, x_N)}{\int \mathrm{d}x_N \, \mu(\widehat{x}, x_N)} \;=\; \frac{\int_{x_{N-1}}^{\infty} \mathrm{d}x_N \, \mathrm{e}^{-N^C x_N^2} \prod_{i<N} (x_N - x_i) \mathrm{e}^{-\frac{N}{4} x_N^2}}{\int_{x_{N-1}}^{\infty} \mathrm{d}x_N \, \prod_{i<N} (x_N - x_i) \mathrm{e}^{-\frac{N}{4} x_N^2}} \;\geqslant\; \mathrm{e}^{-N^C - N^C \sum_{i<N} x_i^2} \tag{4.33}$$

by a straightforward calculation, we get

$$\langle \theta_3 g_t \rangle(\widehat{x}) \;\leqslant\; \mathrm{e}^{N^C + N^C \sum_{i<N} x_i^2}\,.$$

We now derive a lower bound on $\langle \theta_3 g_t \rangle$. Using (4.32) and (4.31) we find

$$\begin{aligned}
\langle \theta_3 g_t \rangle(\widehat{x}) \;&\geqslant\; \mathrm{e}^{-N^C} \frac{\int \mathrm{d}x_N \, \theta_3(x_N) f_t(\widehat{x}, x_N) \mu(\widehat{x}, x_N)}{\int \mathrm{d}x_N \, \mu(\widehat{x}, x_N)} \\
&\geqslant\; \mathrm{e}^{-N^C - N^C \sum_{i<N} x_i^2} \frac{\int_{2+\sigma/2}^{\infty} \mathrm{d}x_N \, \mathrm{e}^{-N^C x_N^2} \mu(\widehat{x}, x_N)}{\int_{x_{N-1}}^{\infty} \mathrm{d}x_N \, \mu(\widehat{x}, x_N)} \\
&\geqslant\; \mathrm{e}^{-N^C - N^C \sum_{i<N} x_i^2}\,,
\end{aligned}$$

by a calculation similar to (4.33). The claim follows from

$$(\theta_1 \theta_2)(\widehat{x}) |\log \langle \theta g_t \rangle (\widehat{x})|^2 \;\leqslant\; 2(\theta_1 \theta_2)(\widehat{x}) |\log \theta_1 \theta_2|^2 + 2(\theta_1 \theta_2)(\widehat{x}) |\log \langle \theta_3 g_t \rangle (\widehat{x})|^2 \;\leqslant\; 2 + N^C\,. \qquad \square$$

5. Bulk universality: proof of Theorem 2.5

Similarly to (2.8), we define $p_{t,N}^{(N-1)}(x_1, \dots, x_{N-1})$ as the probability density obtained by symmetrizing (in the variables $x_1, \dots, x_{N-1}$) the function $\int \mathrm{d}x_N \, f_t(\mathbf{x}) \, \mu(\mathbf{x})$, and set, for $n \leqslant N-1$,

$$p_{t,N}^{(n)}(x_1, \dots, x_n) \;:=\; \int \mathrm{d}x_{n+1} \cdots \mathrm{d}x_{N-1} \, p_{t,N}^{(N-1)}(x_1, \dots, x_{N-1})\,.$$

We begin with a universality result for sparse matrices with a small Gaussian convolution.

THEOREM 5.1. *Let $E \in [-2+\kappa, 2-\kappa]$ for some $\kappa > 0$ and let $b \equiv b_N$ satisfy $|b| \leqslant \kappa/2$. Pick $\varepsilon, \beta > 0$, and set $\tau := N^{-2\alpha+\beta}$, where*

$$\alpha \;\equiv\; \alpha(\phi) \;:=\; \min\biggl\{2\phi - \frac{1}{2}\,,\, 4\phi - \frac{2}{3}\biggr\}\,. \tag{5.1}$$

Let $n \in \mathbb{N}$ and $O : \mathbb{R}^n \to \mathbb{R}$ be compactly supported and continuous. Then there is a $\bar{\tau} \in [\tau/2, \tau]$ such that

$$\biggl|\int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_n \, O(\alpha_1, \dots, \alpha_n) \, \frac{1}{\varrho_{sc}(E)^n} \bigl(p^{(n)}_{\bar{\tau},N} - p^{(n)}_{\mathrm{GOE},N}\bigr)\biggl(E' + \frac{\alpha_1}{N\varrho_{sc}(E)}, \dots, E' + \frac{\alpha_n}{N\varrho_{sc}(E)}\biggr)\biggr| \\ \leqslant\; C_n N^{\varepsilon}\Bigl[b^{-1}N^{-2\phi} + b^{-1/2}N^{-\beta/2}\Bigr]\,. \tag{5.2}$$

PROOF. The claim follows from Theorem 4.2 and Theorem 3.4, similarly to the proof of Theorem 2.1 in [16]. We use that $Q \leqslant (\log N)^{C\xi} N^{-2\alpha}$, as follows from (3.16); the contribution of the low probability complement event to (3.16) may be easily estimated using Cauchy-Schwarz and the estimate $\sum_i \mathbb{E}^t x_i^4 = \mathbb{E}^t \operatorname{Tr} A^4 \leqslant N^C$, uniformly for $t \geqslant 0$. The assumption IV of [16] is a straightforward consequence of the local semicircle law, Theorem 3.3. $\square$

PROPOSITION 5.2. *Let $A^{(1)} = (a^{(1)}_{ij})$ and $A^{(2)} = (a^{(2)}_{ij})$ be sparse random matrices, both satisfying Definition 2.2 with*

$$q^{(1)} \;\sim\; q^{(2)} \;\geqslant\; N^{\phi}$$

(in self-explanatory notation). Suppose that, for each i,j, the first three moments of $a^{(1)}_{ij}$ and $a^{(2)}_{ij}$ are the same, and that the fourth moments satisfy

$$\bigl|\mathbb{E}\bigl(a^{(1)}_{ij}\bigr)^4 - \mathbb{E}\bigl(a^{(2)}_{ij}\bigr)^4\bigr| \;\leqslant\; N^{-2-\delta}\,, \tag{5.3}$$

for some $\delta > 0$.

Let $n \in \mathbb{N}$ and let $F \in C^5(\mathbb{C}^n)$. We assume that, for any multi-index $\alpha \in \mathbb{N}^n$ with $1 \leqslant |\alpha| \leqslant 5$ and any sufficiently small $\varepsilon' > 0$, we have

$$\max\biggl\{\bigl|\partial^\alpha F(x_1, \dots, x_n)\bigr| : \sum_i |x_i| \leqslant N^{\varepsilon'}\biggr\} \;\leqslant\; N^{C_0\varepsilon'}\,, \quad \max\biggl\{\bigl|\partial^\alpha F(x_1, \dots, x_n)\bigr| : \sum_i |x_i| \leqslant N^2\biggr\} \;\leqslant\; N^{C_0}\,,$$

where C_0 is a constant.

Let $\kappa > 0$ be arbitrary. Choose a sequence of positive integers $k_1, \dots, k_n$ and real parameters $E^m_j \in [-2+\kappa, 2-\kappa]$, where $m = 1, \dots, n$ and $j = 1, \dots, k_m$. Let $\varepsilon > 0$ be arbitrary and choose η with $N^{-1-\varepsilon} \leqslant \eta \leqslant N^{-1}$. Set $z^m_j := E^m_j \pm \mathrm{i}\eta$ with an arbitrary choice of the $\pm$ signs.

Then, abbreviating $G^{(l)}(z) := (A^{(l)} - z)^{-1}$, we have

$$\left|\mathbb{E}F\left(\frac{1}{N^{k_1}} \operatorname{Tr}\left[\prod_{j=1}^{k_1} G^{(1)}(z^1_j)\right], \dots, \frac{1}{N^{k_n}} \operatorname{Tr}\left[\prod_{j=1}^{k_n} G^{(1)}(z^n_j)\right]\right) - \mathbb{E}F\bigl(G^{(1)} \to G^{(2)}\bigr)\right| \\ \leqslant\; CN^{1-3\phi+C\varepsilon} + CN^{-\delta+C\varepsilon}\,.$$

Proof. The proof of Theorem 2.3 in [17] may be reproduced almost verbatim; the rest term in the Green function expansion is estimated by an L^∞-L^1 bound using $\mathbb{E}|a_{ij}^{(l)}|^5 \leqslant CN^{-1-3\phi}$. $\square$

As in [17] (Theorem 6.4), Proposition 5.2 readily implies the following correlation function comparison theorem.

Theorem 5.3. *Suppose the assumptions of Proposition 5.2 hold. Let $p_{(1),N}^{(n)}$ and $p_{(2),N}^{(n)}$ be n-point correlation functions of the eigenvalues of $A^{(1)}$ and $A^{(2)}$ respectively. Then for any $|E| < 2$, any $n \geqslant 1$ and any compactly supported test function $O : \mathbb{R}^n \to \mathbb{R}$ we have*

$$\lim_{N\to\infty} \int \mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_n \, O(\alpha_1, \dots, \alpha_n) \Big(p_{(1),N}^{(n)} - p_{(2),N}^{(n)} \Big) \Big(E + \frac{\alpha_1}{N}, \dots, E + \frac{\alpha_n}{N} \Big) \;=\; 0\,.$$

We may now complete the proof of Theorem 2.5.

Proof of Theorem 2.5. In order to invoke Theorems 5.1 and 5.3, we construct a sparse matrix A_0, satisfying Definition 2.2, such that its time evolution $A_{\bar\tau}$ is close to A in the sense of the assumptions of Proposition 5.2. For definiteness, we concentrate on off-diagonal elements (the diagonal elements are dealt with similarly).

For the following we fix $i < j$; all constants in the following are uniform in i, j, and N. Let ξ, ξ', ξ_0 be random variables equal in distribution to $a_{ij}, (a_{\bar\tau})_{ij}, (a_0)_{ij}$ respectively. For any random variable X we use the notation $\widetilde{X} := X - \mathbb{E}X$. Abbreviating $\gamma := 1 - \mathrm{e}^{-\bar\tau}$, we have

$$\xi' \;=\; \sqrt{1-\gamma}\,\xi_0 + \sqrt{\gamma}\, g\,,$$

where g is a centred Gaussian with variance $1/N$, independent of ξ_0. We shall construct a random variable ξ_0, supported on at most three points, such that A_0 satisfies Definition 2.2 and the first four moments of ξ' are sufficiently close to those of ξ. For $k = 1, 2, \dots$ we denote by $m_k(X)$ the k-th moment of a random variable X. We set

$$\xi_0 \;=\; \frac{f}{\sqrt{1-\gamma}\, N} + \widetilde{\xi}_0\,, \tag{5.4}$$

where $m_1(\widetilde{\xi}_0) = 0$ and $m_2(\widetilde{\xi}_0) = N^{-1}$. It is easy to see that $m_k(\xi) = m_k(\xi')$ for $k = 1, 2$.

We take the law of $\widetilde{\xi}_0$ to be of the form

$$p\delta_a + q\delta_{-b} + (1-p-q)\delta_0$$

where $a, b, p, q \geqslant 0$ are parameters satisfying $p + q \leqslant 1$. The conditions $m_1(\widetilde{\xi}_0) = 0$ and $m_2(\widetilde{\xi}_0) = N^{-1}$ imply

$$p \;=\; \frac{1}{aN(a+b)}\,, \qquad q \;=\; \frac{1}{bN(a+b)}\,.$$

Thus, we parametrize ξ_0 using a and b; the condition $p + q \leqslant 1$ reads $ab \geqslant N^{-1}$. Our aim is to determine a and b so that ξ_0 satisfies (2.4), and so that the third and fourth moments of ξ' and ξ are close. By explicit computation we find

$$m_3(\widetilde{\xi}_0) \;=\; \frac{a-b}{N}\,, \qquad m_4(\widetilde{\xi}_0) \;=\; N m_3(\widetilde{\xi}_0)^2 + \frac{ab}{N}\,. \tag{5.5}$$

Now we require that a and b be chosen so that $ab \geqslant N^{-1}$ and

$$m_3(\widetilde{\xi}_0) \;=\; (1-\gamma)^{-3/2} m_3(\widetilde{\xi})\,, \qquad m_4(\widetilde{\xi}_0) \;=\; N m_3(\widetilde{\xi}_0)^2 + m_4(\widetilde{\xi}) - N m_3(\widetilde{\xi})^2\,.$$

Using (5.5), it is easy to see that such a pair (a, b) exists provided that $m_4(\widetilde{\xi}) - N m_3(\widetilde{\xi})^2 \geqslant N^{-2}$. This latter estimate is generally valid for any random variable with $m_1 = 0$; it follows from the elementary inequality $m_4 m_2 - m_3^2 \geqslant m_2^3$ valid whenever $m_1 = 0$.

Next, using (5.5) and the estimates $m_3(\widetilde{\xi}) = O(N^{-1-\phi})$, $m_4(\widetilde{\xi}) = O(N^{-1-2\phi})$, we find

$$a - b \;=\; O(N^{-\phi})\,, \qquad ab \;=\; O(N^{-2\phi})\,,$$

which implies $a, b = O(N^{-\phi})$. We have hence proved that A_0 satisfies Definition 2.2.

One readily finds that $m_3(\xi') = m_3(\xi)$. Moreover, using

$$m_4(\widetilde{\xi}_0) - m_4(\widetilde{\xi}) \;=\; N m_3(\widetilde{\xi})^2 \big[(1-\gamma)^{-3} - 1\big] \;=\; O(N^{-1-2\phi}\gamma)\,,$$

we find

$$m_4(\widetilde{\xi}') - m_4(\widetilde{\xi}) \;=\; (1-\gamma)^2 m_4(\widetilde{\xi}_0) + \frac{6\gamma}{N^2} - \frac{3\gamma^2}{N^2} - m_4(\widetilde{\xi}) \;=\; O(N^{-1-2\phi}\gamma)\,.$$

Summarizing, we have proved

$$m_k(\xi') \;=\; m_k(\xi) \quad (k = 1, 2, 3), \qquad |m_4(\xi') - m_4(\xi)| \;\leqslant\; C N^{-1-2\phi} \bar{\tau}\,. \tag{5.6}$$

The claim follows now by setting $\delta = 2\alpha(\phi) + 2\phi - 1 - \beta$ in (5.3), and invoking Theorems 5.1 and 5.3. $\square$

6. Edge universality: proof of Theorem 2.7

6.1. Rank-one perturbations of the GOE. We begin by deriving a simple, entirely deterministic, result on the eigenvalues of rank-one perturbations of matrices. We choose the perturbation to be proportional to $|\mathbf{e}\rangle\langle\mathbf{e}|$, but all results of this subsection hold trivially if $\mathbf{e}$ is replaced with an arbitrary ℓ^2-normalized vector.

Lemma 6.1 (Monotonicity and interlacing). *Let H be a symmetric $N \times N$ matrix. For $f \geqslant 0$ we set*

$$A(f) \;:=\; H + f|\mathbf{e}\rangle\langle\mathbf{e}|\,.$$

Denote by $\lambda_1 \leqslant \cdots \leqslant \lambda_N$ the eigenvalues of H, and by $\mu_1(f) \leqslant \cdots \leqslant \mu_N(f)$ the eigenvalues of $A(f)$. Then for all $\alpha = 1, \dots, N-1$ and $f \geqslant 0$ the function $\mu_\alpha(f)$ is nondecreasing, satisfies $\mu_\alpha(0) = \lambda_\alpha$, and has the interlacing property

$$\lambda_\alpha \;\leqslant\; \mu_\alpha(f) \;\leqslant\; \lambda_{\alpha+1}\,. \tag{6.1}$$

Proof. From [11], Equation (6.3), we find that μ is an eigenvalue of $H + f|\mathbf{e}\rangle\langle\mathbf{e}|$ if and only if

$$\sum_\alpha \frac{|\langle \mathbf{u}_\alpha\,, \mathbf{e}\rangle|^2}{\mu - \lambda_\alpha} \;=\; \frac{1}{f}\,, \tag{6.2}$$

where $\mathbf{u}_\alpha$ is the eigenvector of H associated with the eigenvalue λ_α. The right-hand side of (6.2) has N singularities at $\lambda_1, \dots, \lambda_N$, away from which it is decreasing. All claims now follow easily. $\square$

Next, we establish the following "eigenvalue sticking" property for GOE. Let α label an eigenvalue close to the right (say) spectral edge. Roughly we prove that, in the case where $H = V$ is a GOE matrix and $f > 1$, the eigenvalue μ_α of $V + f|\mathbf{e}\rangle\langle\mathbf{e}|$ "sticks" to $\lambda_{\alpha+1}$ with a precision $(\log N)^{C\xi} N^{-1}$. This behaviour can be interpreted as a form of long-distance level repulsion, in which the eigenvalues μ_β, $\beta < \alpha$, repel the eigenvalue μ_α and push it close to its maximum possible value, $\lambda_{\alpha+1}$.

Lemma 6.2 (Eigenvalue sticking). *Let V be an $N \times N$ GOE matrix. Suppose moreover that ξ satisfies* (3.10) *and that f satisfies $f \geqslant 1 + \varepsilon_0$. Then there is a $\delta \equiv \delta(\varepsilon_0) > 0$ such that for all α satisfying $N(1-\delta) \leqslant \alpha \leqslant N-1$ we have with (ξ,ν)-high probability*

$$|\lambda_{\alpha+1} - \mu_\alpha| \;\leqslant\; \frac{(\log N)^{C\xi}}{N}\,. \tag{6.3}$$

Similarly, if α instead satisfies $\alpha \leqslant N\delta$ we have with (ξ,ν)-high probability

$$|\mu_\alpha - \lambda_\alpha| \;\leqslant\; \frac{(\log N)^{C\xi}}{N}\,. \tag{6.4}$$

For the proof of Lemma 6.2 we shall need the following result about Wigner matrices, proved in [19].

Lemma 6.3. *Let H be a Wigner matrix with eigenvalues $\lambda_1 \leqslant \cdots \leqslant \lambda_N$ and associated eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_N$. Assume that ξ is given by* (3.1)*. Then the following two statements hold with (ξ,ν)-high probability:*

$$\max_\alpha \|\mathbf{u}_\alpha\|_\infty \;\leqslant\; \frac{(\log N)^{C\xi}}{\sqrt{N}}\,, \tag{6.5}$$

and

$$|\lambda_\alpha - \gamma_\alpha| \;\leqslant\; (\log N)^{C\xi} N^{-2/3} \big(\min\{\alpha, N+1-\alpha\}\big)^{-1/3}\,. \tag{6.6}$$

Moreover, let L satisfy (3.11) *and write $G^H_{ij}(z) := \big[(H-z)^{-1}\big]_{ij}$. Then we have, with (ξ,ν)-high probability,*

$$\bigcap_{z \in D_L} \left\{ \max_{1 \leqslant i,j \leqslant N} \big|G^H_{ij}(z) - \delta_{ij} m_{sc}(z)\big| \leqslant (\log N)^{C\xi} \left(\sqrt{\frac{\operatorname{Im} m_{sc}(z)}{N\eta}} + \frac{1}{N\eta} \right) \right\}, \tag{6.7}$$

where D_L was defined in (3.2)

Proof of Lemma 6.2. We only prove (6.3); the proof of (6.4) is analogous. By orthogonal invariance of V, we may replace $\mathbf{e}$ with the vector $(1,0,\dots,0)$. Let us abbreviate $\zeta_\beta := |u_\beta(1)|^2$. Note that (6.5) implies

$$\max_\beta \zeta_\beta \;\leqslant\; (\log N)^{C\xi} N^{-1} \tag{6.8}$$

with (ξ,ν)-high probability. Now from we (6.2) we get

$$\frac{\zeta_\alpha}{\mu_\alpha - \lambda_{\alpha+1}} + \sum_{\beta \neq \alpha+1} \frac{\zeta_\beta}{\mu_\alpha - \lambda_\beta} \;=\; \frac{1}{f}\,,$$

which yields

$$|\lambda_{\alpha+1} - \mu_\alpha| \;=\; \zeta_\alpha \left| \sum_{\beta \neq \alpha+1} \frac{\zeta_\beta}{\lambda_\beta - \mu_\alpha} + \frac{1}{f} \right|^{-1}. \tag{6.9}$$

We estimate from below, introducing an arbitrary $\eta > 0$,

$$
\begin{aligned}
-\sum_{\beta \neq \alpha+1} \frac{\zeta_\beta}{\lambda_\beta - \mu_\alpha} &= \sum_{\beta < \alpha+1} \frac{\zeta_\beta}{\mu_\alpha - \lambda_\beta} - \sum_{\beta > \alpha+1} \frac{\zeta_\beta}{\lambda_\beta - \mu_\alpha} \\
&\geqslant \sum_{\beta < \alpha+1} \frac{\zeta_\beta(\mu_\alpha - \lambda_\beta)}{(\mu_\alpha - \lambda_\beta)^2 + \eta^2} - \sum_{\beta > \alpha+1} \frac{\zeta_\beta}{\lambda_\beta - \mu_\alpha} \\
&= -\operatorname{Re} G_{11}^V(\mu_\alpha + \mathrm{i}\eta) + \sum_{\beta > \alpha+1} \frac{\zeta_\beta(\lambda_\beta - \mu_\alpha)}{(\lambda_\beta - \mu_\alpha)^2 + \eta^2} - \sum_{\beta > \alpha+1} \frac{\zeta_\beta}{\lambda_\beta - \mu_\alpha} \\
&\geqslant -\operatorname{Re} G_{11}^V(\mu_\alpha + \mathrm{i}\eta) - \sum_{\beta > \alpha+1} \frac{\zeta_\beta \eta^2}{(\lambda_\beta - \mu_\alpha)^3}\,,
\end{aligned} \tag{6.10}
$$

where in the third step we used that $\lambda_{\alpha+1} \geqslant \mu_\alpha$ by (6.1).

We now choose $\eta = (\log N)^{C_1 \log\log N} N^{-1}$. For C_1 large enough, we get from (6.7) that $G_{11}^V(\mu_\alpha + \mathrm{i}\eta) = m_{sc}(\mu_\alpha + \mathrm{i}\eta) + o(1)$. Therefore (3.6) yields

$$
-\operatorname{Re} G_{11}^V(\mu_\alpha + \mathrm{i}\eta) \;\geqslant\; 1 - 2\sqrt{|2 - \mu_\alpha|} + o(1)\,. \tag{6.11}
$$

From (6.6) and (6.1) we get that $|\mu_\alpha - \gamma_\alpha| \leqslant (\log N)^{C\xi} N^{-2/3}$ with (ξ,ν)-high probability. Moreover, the definition (3.15) and $\alpha \geqslant N(1-\delta)$ imply $|\gamma_\alpha - 2| \leqslant C\delta^{2/3}$. Thus we get, with (ξ,ν)-high probability, that $|2 - \mu_\alpha| = o(1) + C\delta^{2/3}$. Therefore (6.11) yields, with (ξ,ν)-high probability,

$$
-\operatorname{Re} G_{11}^V(\mu_\alpha + \mathrm{i}\eta) \;\geqslant\; 1 + o(1) - C\delta^{1/3}\,.
$$

Recalling (6.8), we therefore get from (6.10), with (ξ,ν)-high probability,

$$
\left| \sum_{\beta \neq \alpha+1} \frac{\zeta_\beta}{\lambda_\beta - \mu_\alpha} \right| \;\geqslant\; 1 + o(1) - C\delta^{1/3} - \frac{m(\log N)^{C\xi}}{N^3 |\lambda_{\alpha+1} - \mu_\alpha|^3} - \frac{(\log N)^{C\xi}}{N^3} \sum_{\beta > \alpha+m} \frac{1}{|\lambda_\beta - \mu_\alpha|^3}\,, \tag{6.12}
$$

for any $m \in \mathbb{N}$.

Next, from (6.6) we find that, provided C_2 is large enough, $m := (\log N)^{C_2\xi}$, and $\beta > \alpha + m$, then we have with (ξ,ν)-high probability

$$
|\lambda_\beta - \lambda_{\alpha+1}| \;\geqslant\; |\gamma_\beta - \gamma_{\alpha+1}| - \frac{(\log N)^{C\xi}}{N^{2/3}(N+1-\beta)^{1/3}} \;\geqslant\; c|\gamma_\beta - \gamma_{\alpha+1}|\,.
$$

Then for C_2 large enough we have, with (ξ,ν)-high probability,

$$
\sum_{\beta > \alpha+m} \frac{1}{|\lambda_\beta - \mu_\alpha|^3} \;\leqslant\; C \sum_{\beta > \alpha+m} \frac{1}{(\gamma_\beta - \gamma_{\alpha+1})^3} \;\leqslant\; \frac{CN^3}{(\log N)^{3C_2\xi}}\,.
$$

Thus we get from (6.12), with (ξ,ν)-high probability,

$$
\left| \sum_{\beta \neq \alpha+1} \frac{\zeta_\beta}{\lambda_\beta - \mu_\alpha} \right| \;\geqslant\; 1 + o(1) - C\delta^{1/3} - \frac{(\log N)^{C\xi}}{N^3 |\lambda_{\alpha+1} - \mu_\alpha|^3}\,.
$$

Plugging this into (6.9) and recalling that $f \geqslant 1+\varepsilon_0 > 1$ yields, with (ξ,ν)-high probability,

$$|\lambda_{\alpha+1}-\mu_\alpha| \;\leqslant\; \frac{(\log N)^{C\xi}}{N}\biggl(\varepsilon_0 - C\delta^{1/3} - o(1) - \frac{(\log N)^{C\xi}}{N^3|\lambda_{\alpha+1}-\mu_\alpha|^2}\biggr)^{-1}\,,$$

from which the claim follows. □

6.2. Proof of Theorem 2.7. In this section we prove Theorem 2.7 by establishing the following comparison result for sparse matrices. Throughout the following we shall abbreviate the lower bound in (2.7) by

$$f_* \;:=\; 1+\varepsilon_0\,. \tag{6.13}$$

Proposition 6.4. *Let $\mathbb{P}^{\mathbf{v}}$ and $\mathbb{P}^{\mathbf{w}}$ be laws on the symmetric $N\times N$ random matrices H, each satisfying Definition 2.1 with $q \geqslant N^\phi$ for some ϕ satisfying $1/3 < \phi \leqslant 1/2$. In particular, we have the moment matching condition*

$$\mathbb{E}^{\mathbf{v}} h_{ij} \;=\; \mathbb{E}^{\mathbf{w}} h_{ij} \;=\; 0\,, \qquad \mathbb{E}^{\mathbf{v}} h_{ij}^2 \;=\; \mathbb{E}^{\mathbf{w}} h_{ij}^2 \;=\; \frac{1}{N}\,. \tag{6.14}$$

Set $f := f_*$ in Definition 2.2: $A \equiv A(f_*) = (a_{ij}) := H + f_*|\mathbf{e}\rangle\langle\mathbf{e}|$. As usual, we denote the ordered eigenvalues of H by $\lambda_1 \leqslant \cdots \leqslant \lambda_N$ and the ordered eigenvalues of A by $\mu_1 \leqslant \cdots \leqslant \mu_N$.

Then there is a $\delta > 0$ such that for any $s \in \mathbb{R}$ we have

$$\begin{aligned}&\mathbb{P}^{\mathbf{v}}\Bigl(N^{2/3}(\lambda_N - 2) \leqslant s - N^{-\delta}\Bigr) - N^{-\delta}\\ &\qquad\leqslant\; \mathbb{P}^{\mathbf{w}}\bigl(N^{2/3}(\lambda_N-2) \leqslant s\bigr) \;\leqslant\; \mathbb{P}^{\mathbf{v}}\Bigl(N^{2/3}(\lambda_N-2) \leqslant s + N^{-\delta}\Bigr) + N^{-\delta}\end{aligned} \tag{6.15}$$

as well as

$$\begin{aligned}&\mathbb{P}^{\mathbf{v}}\Bigl(N^{2/3}(\mu_{N-1} - 2) \leqslant s - N^{-\delta}\Bigr) - N^{-\delta}\\ &\qquad\leqslant\; \mathbb{P}^{\mathbf{w}}\bigl(N^{2/3}(\mu_{N-1}-2) \leqslant s\bigr) \;\leqslant\; \mathbb{P}^{\mathbf{v}}\Bigl(N^{2/3}(\mu_{N-1}-2) \leqslant s + N^{-\delta}\Bigr) + N^{-\delta}\end{aligned} \tag{6.16}$$

for $N \geqslant N_0$ sufficiently large, where N_0 is independent of s.

Assuming Proposition 6.4 is proved, we may easily complete the proof of Theorem 2.7 using the results of Section 6.1.

Proof of Theorem 2.7. Choose $\mathbb{P}^{\mathbf{v}}$ to be the law of GOE (see Remark 2.4), and choose $\mathbb{P}^{\mathbf{w}}$ to be the law of a sparse matrix satisfying Definition 2.1 with $q \geqslant N^\phi$. We prove (2.11); the proof of (2.12) is similar.

For the following we write $\mu_\alpha(f) \equiv \mu_\alpha$ to emphasize the f-dependence of the eigenvalues of $A(f)$. Using first (6.1) and then (6.15) we get

$$\mathbb{P}^{\mathbf{w}}\Bigl(N^{2/3}(\mu_{N-1}(f)-2) \leqslant s\Bigr) \;\geqslant\; \mathbb{P}^{\mathbf{w}}\Bigl(N^{2/3}(\lambda_N - 2) \leqslant s\Bigr) \;\geqslant\; \mathbb{P}^{\mathbf{v}}\Bigl(N^{2/3}(\lambda_N-2) \leqslant s - N^{-\delta}\Bigr) - N^{-\delta}\,,$$

for some $\delta > 0$. Next, using first the monotonicity of $\mu_\alpha(f)$ from Lemma 6.1, then (6.16), and finally (6.3), we get

$$\begin{aligned}&\mathbb{P}^{\mathbf{w}}\Bigl(N^{2/3}(\mu_{N-1}(f)-2) \leqslant s\Bigr) \;\leqslant\; \mathbb{P}^{\mathbf{w}}\Bigl(N^{2/3}(\mu_{N-1}(f_*)-2) \leqslant s\Bigr)\\ &\qquad\leqslant\; \mathbb{P}^{\mathbf{v}}\Bigl(N^{2/3}(\mu_{N-1}(f_*)-2) \leqslant s + N^{-\delta}\Bigr) + N^{-\delta} \;\leqslant\; \mathbb{P}^{\mathbf{v}}\Bigl(N^{2/3}(\lambda_N - 2) \leqslant s + 2N^{-\delta}\Bigr) + 2N^{-\delta}\,,\end{aligned}$$

for some $\delta > 0$. This concludes the proof of (2.11), after a renaming of δ. □

The rest of this section is devoted to the proof of Proposition 6.4. We shall only prove (6.16). The proof of (6.15) is similar (in fact easier), and relies on the local semicircle law, Theorem 3.3, with $f = 0$; if $f = 0$ some of the following analysis simplifies (e.g. the proof of Lemma 6.8 below may be completed without the estimate from Lemma 6.9.)

From now on we always assume the setup of Proposition 6.4. In particular, f will always be equal to f_*.

We begin by outlining the proof of Proposition 6.4. The basic strategy is similar to the one used for Wigner matrices in [19] and [25]. For any $E_1 \leqslant E_2$, let

$$\mathcal{N}(E_1, E_2) \;:=\; \bigl|\{\alpha : E_1 \leqslant \mu_\alpha \leqslant E_2\}\bigr|$$

denote the number of eigenvalues of A in the interval $[E_1, E_2]$. In the first step, we express the distribution function in terms of Green functions according to

$$\mathbb{P}^{\mathbf{u}}\bigl(\mu_{N-1} \geqslant E\bigr) \;=\; \mathbb{E}^{\mathbf{u}} K\bigl(\mathcal{N}(E,\infty) - 2\bigr) \;\approx\; \mathbb{E}^{\mathbf{u}} K\bigl(\mathcal{N}(E,E_*) - 1\bigr) \;\approx\; \mathbb{E}^{\mathbf{u}} K\biggl(\int_E^{E_*} \mathrm{d}y \, N \,\mathrm{Im}\, m(y + \mathrm{i}\eta) - 1\biggr). \tag{6.17}$$

Here $\mathbf{u}$ stands for either $\mathbf{v}$ or $\mathbf{w}$, $\eta := N^{-2/3-\varepsilon}$ for some $\varepsilon > 0$ small enough, $K : \mathbb{R} \to \mathbb{R}_+$ is a smooth cutoff function satisfying

$$K(x) = 1 \quad \text{if} \quad |x| \leqslant 1/9 \qquad \text{and} \qquad K(x) = 0 \quad \text{if} \quad |x| \geqslant 2/9\,, \tag{6.18}$$

and

$$E_* \;:=\; 2 + 2(\log N)^{C_0 \xi} N^{-2/3} \tag{6.19}$$

for some C_0 large enough. The first approximate identity in (6.17) follows from Theorem 3.4 which guarantees that $\mu_{N-1} \leqslant E_*$ with (ξ,ν)-high probability, and from (3.21) which guarantees that $\mu_N \geqslant 2 + \sigma$ with (ξ,ν)-high probability. The second approximate identity in (6.17) follows from the approximation

$$\int_{E_1}^{E_2} \mathrm{d}y \, N \,\mathrm{Im}\, m(y + \mathrm{i}\eta) \;=\; \sum_\alpha \int_{E_1}^{E_2} \mathrm{d}y \, \frac{\eta}{(y - \mu_\alpha)^2 + \eta^2} \;\approx\; \mathcal{N}(E_1, E_2)\,,$$

which is valid for E_1 and E_2 near the spectral edge, where the typical eigenvalue separation is $N^{-2/3} \gg \eta$.

The second step of our proof is to compare expressions such as the right-hand side of (6.17) for $\mathbf{u} = \mathbf{v}$ and $\mathbf{u} = \mathbf{w}$. This is done using a Lindeberg replacement strategy and a resolvent expansion of the argument of K. This step is implemented in Section 6.3, to which we also refer for a heuristic discussion of the argument.

Now we give the rigorous proof of the steps outlined in (6.17). We first collect the tools we shall need. From (3.18) and (3.21) we get that there is a constant $C_0 > 0$ such that, under both $\mathbb{P}^{\mathbf{v}}$ and $\mathbb{P}^{\mathbf{w}}$, we have with (ξ,ν)-high probability

$$|N^{2/3}(\mu_{N-1} - 2)| \;\leqslant\; (\log N)^{C_0\xi}\,, \qquad \mu_N \;\geqslant\; 2 + \sigma\,, \tag{6.20}$$

and

$$\mathcal{N}\left(2 - \frac{2(\log N)^{C_0\xi}}{N^{2/3}},\; 2 + \frac{2(\log N)^{C_0\xi}}{N^{2/3}}\right) \;\leqslant\; (\log N)^{2C_0\xi}\,. \tag{6.21}$$

Therefore in (6.16) we can assume that s satisfies

$$-\,(\log N)^{C_0\xi} \;\leqslant\; s \;\leqslant\; (\log N)^{C_0\xi}. \tag{6.22}$$

Recall the definition (6.19) of E_* and and introduce, for any $E \leqslant E_*$, the characteristic function on the interval $[E, E_*]$,

$$\chi_E \;:=\; \mathbf{1}_{[E,E_*]}\,.$$

For any $\eta > 0$ we define

$$\theta_\eta(x) \;:=\; \frac{\eta}{\pi(x^2+\eta^2)} \;=\; \frac{1}{\pi}\operatorname{Im}\frac{1}{x-\mathrm{i}\eta} \tag{6.23}$$

to be an approximate delta function on scale η.

The following result allows us to replace the sharp counting function $\mathcal{N}(E, E_*) = \operatorname{Tr}\chi_E(H)$ with its approximation smoothed on scale η.

Lemma 6.5. *Suppose that E satisfies*

$$|E-2|N^{2/3} \;\leqslant\; (\log N)^{C_0\xi}. \tag{6.24}$$

Let $\ell := \frac{1}{2}N^{-2/3-\varepsilon}$ and $\eta := N^{-2/3-9\varepsilon}$, and recall the definition of the function K from (6.18). Then the following statements hold for both ensembles $\mathbb{P}^{\mathbf{v}}$ and $\mathbb{P}^{\mathbf{w}}$. For some $\varepsilon > 0$ small enough the inequalities

$$\operatorname{Tr}(\chi_{E+\ell} * \theta_\eta)(H) - N^{-\varepsilon} \;\leqslant\; \mathcal{N}(E,\infty) - 1 \;\leqslant\; \operatorname{Tr}(\chi_{E-\ell} * \theta_\eta)(H) + N^{-\varepsilon} \tag{6.25}$$

hold with (ξ,ν)-high probability. Furthermore, we have

$$\mathbb{E}\, K\Big(\operatorname{Tr}(\chi_{E-\ell} * \theta_\eta)(H)\Big) \;\leqslant\; \mathbb{P}(\mathcal{N}(E,\infty)=1) \;\leqslant\; \mathbb{E}\, K\Big(\operatorname{Tr}(\chi_{E+\ell} * \theta_\eta)(H)\Big) + \mathrm{e}^{-\nu(\log N)^\xi} \tag{6.26}$$

for sufficiently large N independent of E, as long as (6.24) holds.

Proof. The proof of Corollary 6.2 in [19] can be reproduced almost verbatim. In the estimate (6.17) of [19], we need the bound, with (ξ,ν)-high probability,

$$\big|m(E+\mathrm{i}\ell) - m_{sc}(E+\mathrm{i}\ell)\big| \;\leqslant\; \frac{(\log N)^{C\xi}}{N\ell}$$

for $N^{-1+c} \leqslant \ell \leqslant N^{-2/3}$. This is an easy consequence of the local semicircle law (3.12) and the assumption $q \geqslant N^{1/3}$.

Note that, when compared to Corollary 6.2 in [19], the quantity $\mathcal{N}(E,\infty)$ has been incremented by one; the culprit is the single eigenvalue $\mu_N \geqslant 2+\sigma$. □

Recalling that $\theta_\eta(H) = \frac{1}{\pi}\operatorname{Im} G(\mathrm{i}\eta)$, Lemma 6.5 bounds the probability of $\mathcal{N}(E,\infty) = 1$ in terms of expectations of functionals of Green functions. We now show that the difference between the expectations of these functionals, with respect to the two probability distributions $\mathbb{P}^{\mathbf{v}}$ and $\mathbb{P}^{\mathbf{w}}$, is negligible assuming their associated second moments of h_{ij} coincide. The precise statement is the following Green function comparison theorem at the edge. All statements are formulated for the upper spectral edge 2, but with the same proof they hold for the lower spectral edge -2 as well.

For the following it is convenient to introduce the shorthand

$$I_\varepsilon \;:=\; \{x : |x-2| \leqslant N^{-2/3+\varepsilon}\} \tag{6.27}$$

where $\varepsilon > 0$.

Proposition 6.6 (Green function comparison theorem on the edge). *Suppose that the assumptions of Proposition 6.4 hold. Let $F:\mathbb{R}\to\mathbb{R}$ be a function whose derivatives satisfy*

$$\sup_x |F^{(n)}(x)|(1+|x|)^{-C_1} \;\leqslant\; C_1 \qquad \textit{for} \quad n \;=\; 1,2,3,4\,, \tag{6.28}$$

with some constant $C_1>0$. Then there exists a constant $\widetilde{\varepsilon}>0$, depending only on C_1, such that for any $\varepsilon<\widetilde{\varepsilon}$ and for any real numbers $E,E_1,E_2\in I_\varepsilon$, and setting $\eta := N^{-2/3-\varepsilon}$, we have

$$\left|\mathbb{E}^{\mathbf{v}} F\left(N\eta \operatorname{Im} m(z)\right) - \mathbb{E}^{\mathbf{w}} F\left(N\eta \operatorname{Im} m(z)\right)\right| \;\leqslant\; C N^{1/3+C\varepsilon} q^{-1} \qquad \textit{for} \quad z \;=\; E+\mathrm{i}\eta\,, \tag{6.29}$$

and

$$\left|\mathbb{E}^{\mathbf{v}} F\left(N\int_{E_1}^{E_2} \mathrm{d}y \;\operatorname{Im} m(y+\mathrm{i}\eta)\right) - \mathbb{E}^{\mathbf{w}} F\left(N\int_{E_1}^{E_2} \mathrm{d}y \;\operatorname{Im} m(y+\mathrm{i}\eta)\right)\right| \;\leqslant\; N^{1/3+C\varepsilon} q^{-1} \tag{6.30}$$

for some constant C and large enough N.

We postpone the proof of Proposition 6.6 to the next section. Assuming it proved, we now have all the ingredients needed to complete the proof of Proposition 6.4.

Proof of Proposition 6.4. As observed after (6.20) and (6.21), we may assume that (6.22) holds. We define $E:=2+sN^{-2/3}$ that satisfies (6.24). We define E_* as in (6.19) with the C_0 such that (6.20) and (6.21) hold. From (6.26) we get, for any sufficiently small $\varepsilon>0$,

$$\mathbb{E}^{\mathbf{w}}\, K\Big(\operatorname{Tr}(\chi_{E-\ell}*\theta_\eta)(H)\Big) \;\leqslant\; \mathbb{P}^{\mathbf{w}}(\mathcal{N}(E,\infty)=1) \tag{6.31}$$

where we set

$$\ell \;:=\; \frac{1}{2}N^{-2/3-\varepsilon}\,, \qquad \eta \;:=\; N^{-2/3-9\varepsilon}.$$

Now (6.30) applied to the case $E_1=E-\ell$ and $E_2=E_*$ shows that there exists a $\delta>0$ such that for sufficiently small $\varepsilon>0$ we have

$$\mathbb{E}^{\mathbf{v}} K\Big(\operatorname{Tr}(\chi_{E-\ell}*\theta_\eta)(H)\Big) \;\leqslant\; \mathbb{E}^{\mathbf{w}}\, K\Big(\operatorname{Tr}(\chi_{E-\ell}*\theta_\eta)(H)\Big) + N^{-\delta} \tag{6.32}$$

(note that here 9ε plays the role of ε in the Proposition 6.6). Next, the second bound of (6.26) yields

$$\mathbb{P}^{\mathbf{v}}(\mathcal{N}(E-2\ell,\infty)=1) \;\leqslant\; \mathbb{E}^{\mathbf{v}}\, K\Big(\operatorname{Tr}(\chi_{E-\ell}*\theta_\eta)(H)\Big) + \mathrm{e}^{-\nu(\log N)^\xi} \tag{6.33}$$

Combining these inequalities, we have

$$\mathbb{P}^{\mathbf{v}}(\mathcal{N}(E-2\ell,\infty)=1) \;\leqslant\; \mathbb{P}^{\mathbf{w}}(\mathcal{N}(E,\infty)=1) + 2N^{-\delta} \tag{6.34}$$

for sufficiently small $\varepsilon>0$ and sufficiently large N. Setting $E=2+sN^{-2/3}$ proves the first inequality of (6.16). Switching the roles of $\mathbf{v}$ and $\mathbf{w}$ in (6.34) yields the second inequality of (6.16). $\square$

6.3. Proof of Proposition 6.6. All that remains is the proof of Proposition 6.6, to which this section is devoted. Throughout this section we suppose that the assumptions of Proposition 6.4 hold, and in particular that $f = 1 + \varepsilon_0$.

We now set up notations to replace the matrix elements one by one. This step is identical for the proof of both (6.29) and (6.30); we use the notations of the case (6.29), for which they are less involved.

For the following it is convenient to slightly modify our notation. We take two copies of our probability space, one of which carries the law $\mathbb{P}^{\mathbf{v}}$ and the other the law $\mathbb{P}^{\mathbf{w}}$. We work on the product space and write $H^{\mathbf{v}}$ for the copy carrying the law $\mathbb{P}^{\mathbf{v}}$ and $H^{\mathbf{w}}$ for the copy carrying the law $\mathbb{P}^{\mathbf{w}}$. The matrices $A^{\mathbf{v}}$ and $A^{\mathbf{w}}$ are defined in the obvious way, and we use the notations $A^{\mathbf{v}} = (v_{ij})$ and $A^{\mathbf{w}} = (w_{ij})$ for their entries. Similarly, we denote by $G^{\mathbf{v}}(z)$ and $G^{\mathbf{w}}(z)$ the Green functions of the matrices $A^{\mathbf{v}}$ and $A^{\mathbf{w}}$.

Fix a bijective ordering map on the index set of the independent matrix elements,

$$\phi : \{(i,j) : 1 \leqslant i \leqslant j \leqslant N\} \;\to\; \Big\{0, \dots, \gamma_{\max}\Big\} \qquad \text{where} \quad \gamma_{\max} \;:=\; \frac{N(N+1)}{2} - 1\,,$$

and denote by A_γ the generalized Wigner matrix whose matrix elements a_{ij} follow the v-distribution if $\phi(i,j) \leqslant \gamma$ and they follow the w-distribution otherwise; in particular $A_0 = A^{\mathbf{v}}$ and $A_{\gamma_{\max}} = A^{\mathbf{w}}$.

Next, set $\eta := N^{-2/3-\varepsilon}$. We use the identity

$$\operatorname{Im} m_{sc}(E + \mathrm{i}\eta) \;\leqslant\; \sqrt{|E-2| + \eta} \;\leqslant\; CN^{-1/3+\varepsilon/2}\,. \tag{6.35}$$

Therefore Theorem 2.9 of [11] yields, with (ξ,ν)-high probability,

$$\max_{0 \leqslant \gamma \leqslant \gamma_{\max}} \max_{1 \leqslant k,l \leqslant N} \max_{E \in I_\varepsilon} \left| \left(\frac{1}{A_\gamma - E - \mathrm{i}\eta} \right)_{kl} - \delta_{kl} m_{sc}(E + \mathrm{i}\eta) \right| \;\leqslant\; \frac{1}{p} \tag{6.36}$$

where we defined

$$\frac{1}{p} \;:=\; N^{\varepsilon} \left(q^{-1} + \frac{1}{N\eta} \right) \;\leqslant\; N^{-1/3+2\varepsilon}\,. \tag{6.37}$$

We set $z = E + \mathrm{i}\eta$ where $E \in I_\varepsilon$ and $\eta = N^{-2/3-\varepsilon}$. Using (6.36), (6.37), and the identity

$$\operatorname{Im} m \;=\; \frac{1}{N} \operatorname{Im} \operatorname{Tr} G \;=\; \frac{\eta}{N} \sum_{ij} G_{ij} \overline{G_{ij}}\,,$$

we find, as in (6.36) of [19], that in order to prove (6.29) it is enough to prove

$$\left| \mathbb{E} F \left(\eta^2 \sum_{i \neq j} G^{\mathbf{v}}_{ij} \overline{G^{\mathbf{v}}_{ji}} \right) - \mathbb{E} F \left(G^{\mathbf{v}} \to G^{\mathbf{w}} \right) \right| \;\leqslant\; CN^{1/3+C\varepsilon} q^{-1} \tag{6.38}$$

at $z = E + \mathrm{i}\eta$. We write the quantity in the absolute value on the left-hand side of (6.38) as a telescopic sum,

$$\begin{aligned} &\mathbb{E}\, F \left(\eta^2 \sum_{i \neq j} \left(\frac{1}{A^{\mathbf{v}} - z} \right)_{ij} \overline{\left(\frac{1}{A^{\mathbf{v}} - z} \right)_{ji}} \right) - \mathbb{E}\, F \left(A^{\mathbf{v}} \to A^{\mathbf{w}} \right) \\ &\qquad\qquad = \; - \sum_{\gamma=2}^{\gamma_{\max}} \Big(\mathbb{E}\, F(A^{\mathbf{v}} \to A_\gamma) - \mathbb{E}\, F(A^{\mathbf{v}} \to A_{\gamma-1}) \Big)\,. \end{aligned} \tag{6.39}$$

Let $E^{(ij)}$ denote the matrix whose matrix elements are zero everywhere except at the (i,j) position, where it is 1, i.e. $E^{(ij)}_{kl} := \delta_{ik}\delta_{jl}$. Fix $\gamma \geqslant 1$ and let (b,d) be determined by $\phi(b,d) = \gamma$. For definiteness, we assume the off-diagonal case $b \neq d$; the case $b = d$ can be treated similarly. Note that the number of diagonal terms is N and the number of off-diagonal terms is $O(N^2)$. We shall compare $A_{\gamma-1}$ with A_γ for each γ and then sum up the differences in (6.39).

Note that these two matrices differ only in the entries (b,d) and (b,d), and they can be written as

$$A_{\gamma-1} \;=\; Q + V \qquad \text{where} \quad V \;:=\; (v_{bd} - \mathbb{E}v_{bd})E^{(bd)} + (v_{db} - \mathbb{E}v_{db})E^{(db)}\,, \tag{6.40}$$

and

$$A_\gamma \;=\; Q + W \qquad \text{where} \quad W \;:=\; (w_{bd} - \mathbb{E}w_{bd})E^{(bd)} + (w_{db} - \mathbb{E}w_{db})E^{(db)}\,,$$

where the matrix Q satisfies

$$Q_{bd} \;=\; Q_{db} \;=\; f/N \;=\; \mathbb{E}v_{bd} \;=\; \mathbb{E}v_{db} \;=\; \mathbb{E}w_{bd} \;=\; \mathbb{E}w_{db}\,,$$

where, we recall $f = 1 + \varepsilon_0$. It is easy to see that

$$\max_{i,j} |v_{ij}| + \max_{i,j} |w_{ij}| \;\leqslant\; (\log N)^{C\xi} q^{-1} \tag{6.41}$$

with (ξ,ν)-high probability, and that

$$\mathbb{E}v_{ij} \;=\; \mathbb{E}w_{ij} \;=\; 0\,, \qquad \mathbb{E}(v_{ij})^2 \;=\; \mathbb{E}(w_{ij})^2 \;\leqslant\; C/N\,, \qquad \mathbb{E}|v_{ij}|^k + \mathbb{E}|w_{ij}|^k \;\leqslant\; CN^{-1}q^{2-k} \tag{6.42}$$

for $k = 2,3,4,5,6$.

We define the Green functions

$$R \;:=\; \frac{1}{Q - z}\,, \qquad S \;:=\; \frac{1}{A_{\gamma-1} - z}\,, \qquad T \;:=\; \frac{1}{A_\gamma - z}\,. \tag{6.43}$$

We now claim that the estimate (6.36) holds for the Green function R as well, i.e.

$$\max_{1\leqslant k,l\leqslant N} \max_{E \in I_\varepsilon} \big| R_{kl}(E + \mathrm{i}\eta) - \delta_{kl} m_{sc}(E + \mathrm{i}\eta)\big| \;\leqslant\; p^{-1} \tag{6.44}$$

holds with (ξ,ν)-high probability. To see this, we use the resolvent expansion

$$R \;=\; S + SVS + (SV)^2 S + \ldots + (SV)^9 S + (SV)^{10} R. \tag{6.45}$$

Since V has only at most two nonzero elements, when computing the entry (k,l) of this matrix identity, each term is a sum of finitely many terms (i.e. the number of summands is independent of N) that involve matrix elements of S or R and v_{ij}, e.g. of the form $(SVS)_{kl} = S_{ki}v_{ij}S_{jl} + S_{kj}v_{ji}S_{il}$. Using the bound (6.36) for the S matrix elements, the bound (6.41) for v_{ij} and the trivial bound $|R_{ij}| \leqslant \eta^{-1} \leqslant N$, we get (6.44).

Having introduced these notations, we may now give an outline of the proof of Proposition 6.6. We have to estimate each summand of the telescopic sum (6.39) with $b \neq d$ (the generic case) by $o(N^{-2})$; in the non-generic case $b = d$, a bound of size $o(N^{-1})$ suffices. For simplicity, assume that we are in the generic case $b \neq d$ and that F has only one argument. Fix $z = E + \mathrm{i}\eta$, where $E \in I_\varepsilon$ (see (6.27)) and $\eta := N^{-2/3-\varepsilon}$. Define

$$y^S \;:=\; \eta^2 \sum_{i \neq j} S_{ij}(z)\overline{S_{ji}(z)}\,; \tag{6.46}$$

the random variable y^R is defined similarly. We shall show that

$$\mathbb{E}F(y^S) \;=\; B + \mathbb{E}F(y^R) + O(N^{-1/3+C\varepsilon}p^{-4}q^{-1})\,, \tag{6.47}$$

for some deterministic B which depends only on the law of Q and the first two moments of v_{bd}. From (6.47) we immediately conclude that (6.29) holds. In order to prove (6.47), we expand

$$F(y^S) - F(y^R) \;=\; F'(y^R)(y^S - y^R) + \frac{1}{2}F''(y^R)(y^S - y^R)^2 + \frac{1}{6}F'''(\zeta)(y^S - y^R)^3\,, \tag{6.48}$$

where ζ lies between y^S and y^R. Next, we apply the resolvent expansion

$$S \;=\; R + RVR + (RV)^2R + \ldots + (RV)^m R + (RV)^{m+1}S \tag{6.49}$$

to each factor S in (6.48) for some $m \geqslant 2$. Here we only concentrate on the linear term in (6.48). The second term is dealt with similarly. (The rest term in (6.48) requires a different treatment because $F'''(\zeta)$ is not independent of v_{bd}. It may however by estimated cheaply using a naive power counting.) By definition, Q is independent of v_{bd}, and hence $F'(y^R)$ and R are independent of h_{bd}. Therefore the expectations of the first and second order terms (in the variable v_{bd}) in $\mathbb{E}F'(y^R)(y^S - y^R)$ are put into B. The third order terms in $\mathbb{E}F'(y^R)(y^S - y^R)$ are bounded, using a naive power counting, by

$$\eta^2 N^2 p^{-3}\mathbb{E}|v_{bd}|^3 \;\leqslant\; N^{-4/3}N^2p^{-3}N^{-1}q^{-1}\,. \tag{6.50}$$

Here we used that, thanks to the assumption $i \neq j$, in the generic terms $\{i,j\} \cap \{b,d\} = \emptyset$ there are at least three off-diagonal matrix elements R in the resolvent expansion of (6.46). Indeed, since $b, d \notin \{i,j\}$, the terms of order greater than one in (6.49) have at least two off-diagonal resolvents matrix elements, and other factor in (6.46) has at least one off-diagonal resolvent matrix element since $i \neq j$. Thus we get a factor p^{-3} by (6.36) (the non-generic terms are suppressed by a factor N^{-1}). Note that the bound (6.50) is still too large compared to N^2, since $p \geqslant N^{-1/3}$. The key observation to solve this problem is that the expectation of the leading term is much smaller than its typical size; this allows us to gain an additional factor p^{-1}. A similar observation was used in [19], but in the present case this estimate (performed in Lemma 6.8 below) is substantially complicated by the non-vanishing expectation of the entries of A. Much of the heavy notation in the following argument arises from the need to keep track of the non-generic terms, which have fewer off-diagonal elements than the generic terms, but have a smaller entropy factor. The improved bound on the difference $\mathbb{E}F'(y^R)(y^S - y^R)$ is

$$N^{-4/3}N^2p^{-4}N^{-1}q^{-1} \;=\; N^{-1/3}p^{-4}q^{-1}\,,$$

which is much smaller than N^{-2} provided that $q \geqslant N^{\phi}$ for $\phi > 1/3$ and ε is small enough.

The key step to the proof of Proposition 6.6 is the following lemma.

Lemma 6.7. *Fix an index $\gamma = \phi(b,d)$ and recall the definitions of Q, R and S from (6.43). For any small enough $\varepsilon > 0$ and under the assumptions in Proposition 6.6, there exists a constant C depending on F (but independent of γ) and constants B_N and D_N, depending on the law $\mathrm{law}(Q)$ of the Green function Q and on the second moments $m_2(v_{bd})$ of v_{bd}, such that, for large enough N (independent of γ) we have*

$$\left|\mathbb{E}F\left(\eta\int_{E_1}^{E_2}\mathrm{d}y\sum_{i\neq j}S_{ij}\overline{S}_{ji}(y+\mathrm{i}\eta)\right) - \mathbb{E}F\left(\eta\int_{E_1}^{E_2}\mathrm{d}y\sum_{i\neq j}R_{ij}\overline{R}_{ji}(y+\mathrm{i}\eta)\right) - B_N\big(m_2(v_{bd}),\mathrm{law}(Q)\big)\right| \\ \leqslant\; N^{\mathbf{1}(b=d)-5/3+C\varepsilon}q^{-1}\,, \tag{6.51}$$

where, we recall, $\eta = N^{-2/3-\varepsilon}$, as well as

$$\left|\mathbb{E}\, F\left(\eta^2 \sum_{i\neq j} S_{ij}\overline{S}_{ji}(z)\right) - \mathbb{E}\, F\left(\eta^2 \sum_{i\neq j} R_{ij}\overline{R}_{ji}(z)\right) - D_N\big(m_2(v_{bd}), \mathrm{law}(Q)\big)\right| \;\leqslant\; N^{\mathbf{1}(b=d)-5/3+C\varepsilon} q^{-1}\,, \tag{6.52}$$

where $z = E + \mathrm{i}\eta$. The constants B_N and D_N may also depend on F, but they depend on the centered random variable v_{bd} only through its second moments.

Assuming Lemma 6.7, we now complete the proof of Proposition 6.6.

Proof of Proposition 6.6. Clearly, Lemma 6.7 also holds if S is replaced by T. Since Q is independent of v_{bd} and w_{bd}, and $m_2(v_{bd}) = m_2(w_{bd}) = 1/N$, we have $D_N\big(m_2(v_{bd}), \mathrm{law}(Q)\big) = D_N\big(m_2(w_{bd}), \mathrm{law}(Q)\big)$. Thus we get from Lemma 6.7 that

$$\left|\mathbb{E}\, F\left(\eta^2 \sum_{i\neq j} S_{ij}\overline{S}_{ji}(z)\right) - \mathbb{E}\, F\left(\eta^2 \sum_{i\neq j} T_{ij}\overline{T}_{ji}(z)\right)\right| \leqslant C N^{\mathbf{1}(b=d)-5/3+C\varepsilon} q^{-1}. \tag{6.53}$$

Recalling the definitions of S and T from (6.43), the bound (6.53) compares the expectation of a function of the resolvent of A_γ and that of $A_{\gamma-1}$. The telescopic summation in (6.39) then implies (6.38), since the number of summands with $b \neq d$ is of order N^2 but the number of summands with $b = d$ is only N. Similarly, (6.51) implies (6.30). This completes the proof. $\square$

Proof of Lemma 6.7. Throughout the proof we abbreviate $A_{\gamma-1} = A = (a_{ij})$ where $a_{ij} = h_{ij} + f/N$. We shall only prove the more complicated case (6.51); the proof of (6.52) is similar. In fact, we shall prove the bound

$$\left|\mathbb{E}F\left(\eta \int_{E_1}^{E_2} \mathrm{d}y \sum_{i\neq j} S_{ij}\overline{S}_{ji}(y+\mathrm{i}\eta)\right) - \mathbb{E}F\left(\eta \int_{E_1}^{E_2} \mathrm{d}y \sum_{i\neq j} R_{ij}\overline{R}_{ji}(y+\mathrm{i}\eta)\right) - B_N\big(m_2(h_{bd}), \mathrm{law}(Q)\big)\right| \\ \leqslant\; C N^{\mathbf{1}(b=d)-1/3+C\varepsilon} p^{-4} q^{-1}\,, \tag{6.54}$$

from which (6.51) follows by (6.37).

From (6.36) we get

$$\max_{1\leqslant k,l\leqslant N}\, \max_{E\in I_\varepsilon} \big|S_{kl}(E+\mathrm{i}\eta) - \delta_{kl} m_{sc}(E+\mathrm{i}\eta)\big| \;\leqslant\; p^{-1} \tag{6.55}$$

with (ξ,ν)-high probability. Define Ω as the event on which (6.55), (6.44), and (6.41) hold. We have proved that Ω holds with (ξ,ν)-high probability. Since the arguments of F in (6.54) are bounded by $CN^{2+2\varepsilon}$ and $F(x)$ increases at most polynomially, it is easy to see that the contribution of the event Ω^c to the expectations in (6.54) is negligible.

Define x^S and x^R by

$$x^S \;:=\; \eta \int_{E_1}^{E_2} \mathrm{d}y \sum_{i\neq j} S_{ij}\overline{S}_{ji}(y+\mathrm{i}\eta)\,, \qquad x^R \;:=\; \eta \int_{E_1}^{E_2} \mathrm{d}y \sum_{i\neq j} R_{ij}\overline{R}_{ji}(y+\mathrm{i}\eta), \tag{6.56}$$

and decompose x^S into three parts

$$x^S \;=\; x_2^S + x_1^S + x_0^S \qquad \text{where} \quad x_k^S \;:=\; \eta \int_{E_1}^{E_2} \mathrm{d}y \sum_{i \neq j} \mathbf{1}\big(|\{i,j\} \cap \{b,d\}| = k\big)\, S_{ij}\overline{S}_{ji}(y + \mathrm{i}\eta)\,; \tag{6.57}$$

x_k^R is defined similarly. Here $k = |\{i,j\} \cap \{b,d\}|$ is the number of times the indices b and d appear among the summation indices i,j. Clearly $k = 0$, 1 or 2. The number of the terms in the sum of the definition of x_k^S is $O(N^{2-k})$. A resolvent expansion yields

$$S \;=\; R - RVR + (RV)^2R - (RV)^3R + (RV)^4R - (RV)^5R + (RV)^6S\,. \tag{6.58}$$

In the following formulas we shall, as usual, omit the spectral parameter from the notation of the resolvents. The spectral parameter is always $y + \mathrm{i}\eta$ with $y \in [E_1, E_2]$; in particular, $y \in I_\varepsilon$.

If $|\{i,j\} \cap \{b,d\}| = k$, recalling that $i \neq j$ we find that there are at least $2 - k$ off-diagonal resolvent elements in $\big[(RV)^m R\big]_{ij}$, so that (6.36) yields in Ω

$$\big|\big[(RV)^m R\big]_{ij}\big| \;\leqslant\; C_m \left(N^\varepsilon q^{-1}\right)^m p^{-(2-k)} \qquad \text{where} \quad m \in \mathbb{N}_+\,, \quad m \leqslant 6\,, \quad k = 0,1,2\,. \tag{6.59}$$

Similarly, we have in Ω

$$\big|\big[(RV)^m S\big]_{ij}\big| \;\leqslant\; C_m \left(N^\varepsilon q^{-1}\right)^m p^{-(2-k)} \qquad \text{where} \quad m \in \mathbb{N}_+\,, \quad m \leqslant 6\,, \quad k = 0,1,2\,. \tag{6.60}$$

Therefore we have in Ω that

$$|x_k^S - x_k^R| \;\leqslant\; CN^{2/3-k}p^{-(3-k)}N^\varepsilon q^{-1} \qquad \text{for} \quad k = 0,1,2\,, \tag{6.61}$$

where the factor $N^{2/3-k}$ comes from $\sum_{i \neq j}$, η and $\int \mathrm{d}E$. Inserting these bounds into the Taylor expansion of F, using

$$q \;\geqslant\; N^\phi \;\geqslant\; N^{1/3+C\varepsilon} \;\geqslant\; p \;\geqslant\; N^{1/3-2\varepsilon} \tag{6.62}$$

and keeping only the terms larger than $O(N^{-1/3+C\varepsilon}p^{-4}q^{-1})$, we obtain

$$\left|\mathbb{E}\big(F(x^S) - F(x^R)\big) - \mathbb{E}\left(F'(x^R)(x_0^S - x_0^R) + \frac{1}{2}F''(x^R)(x_0^S - x_0^R)^2 + F'(x^R)(x_1^S - x_1^R)\right)\right| \\ \leqslant\; CN^{-1/3+C\varepsilon}p^{-4}q^{-1}\,, \tag{6.63}$$

where we used the remark after (6.55) to treat the contribution on the event Ω. Since there is no x_2 appearing in (6.63), we can focus on the cases $k = 0$ and $k = 1$.

To streamline the notation, we introduce

$$R_{ij}^{(m)} \;:=\; (-1)^m \big[(RV)^m R\big]_{ij}\,. \tag{6.64}$$

Then using (6.59) and the estimate $\max_{i \neq j} |R_{ij}| \leqslant p^{-1}$ we get

$$\big|R_{ij}^{(m)}\big| \;\leqslant\; C_m \left(N^\varepsilon q^{-1}\right)^m p^{-(2-k)+\delta_{0m}\delta_{0k}}\,. \tag{6.65}$$

Now we decompose the sum $x_k^S - x_k^R$ according to the number of matrix elements h_{bd} and h_{db}. To that end, for $k \in \{0,1\}$ and $s,t \in \{0,1,2,3,4\}$ and $s+t \geqslant 1$, we define

$$Q_k^{(s,t)} \;:=\; \eta \int_{E_1}^{E_2} \mathrm{d}y \sum_{i \neq j} \mathbf{1}\big(|\{i,j\} \cap \{b,d\}| = k\big)\, R_{ij}^{(s)} \overline{R_{ji}^{(t)}}\,, \tag{6.66}$$

and set

$$Q_k^{(\ell)} \;:=\; \sum_{s+t=\ell} Q_k^{(s,t)}\,. \tag{6.67}$$

Using (6.65) we get the estimates, valid on Ω,

$$\big|Q_k^{(s,t)}\big| \;\leqslant\; C_{st}\left(N^{\varepsilon} q^{-1}\right)^{s+t} N^{2/3-k} p^{-(4-2k)+\delta_{0s}\delta_{0k}+\delta_{0t}\delta_{0k}}\,, \qquad \big|Q_k^{(\ell)}\big| \;\leqslant\; C_\ell \left(N^{\varepsilon} q^{-1}\right)^{\ell} N^{2/3-k} p^{-(3-k)}\,, \tag{6.68}$$

where $\ell \geqslant 1$. Using (6.62), (6.59), and (6.60), we find the decomposition

$$x_k^S - x_k^R \;=\; \sum_{1 \leqslant s+t \leqslant 4} Q_k^{(s,t)} + O(N^{-1/3+C\varepsilon} p^{-4} q^{-1})\,, \tag{6.69}$$

where s and t are non-negative. By (6.68) and (6.44) we have for $s+t \geqslant 1$

$$\left|\mathbb{E}_{bd}\, Q_k^{(s,t)}\right| \;\leqslant\; C_{st} q^{2-s-t} N^{-1/3-k} p^{-(4-2k)+\delta_{0s}\delta_{0k}+\delta_{0t}\delta_{0k}}\,, \tag{6.70}$$

where $\mathbb{E}_{bd}$ denotes partial expectation with respect to the variable h_{bd}. Here we used that only terms with at least two elements h_{bd} or h_{db} survive. Recalling (6.42), we find that taking the partial expectation $\mathbb{E}_{bd}$ improves the bound (6.68) by a factor q^2/N. Thus we also have

$$\left|\mathbb{E}_{bd}\, Q_k^{(\ell)}\right| \leqslant C_\ell\, q^{2-\ell} N^{-1/3-k} p^{-(3-k)} \tag{6.71}$$

Similarly, for $s+t \geqslant 1$ and $u+v \geqslant 1$ we have

$$\left|\mathbb{E}_{bd}\, Q_k^{(s,t)} Q_k^{(u,v)}\right| \;\leqslant\; q^{2-s-t-u-v} N^{1/3-2k} p^{-(8-4k)+\delta_{0s}\delta_{0k}+\delta_{0t}\delta_{0k}+\delta_{0u}\delta_{0k}+\delta_{0v}\delta_{0k}}\,, \tag{6.72}$$

which implies

$$\left|\mathbb{E}_{bd}\, Q_k^{(\ell_1)} Q_k^{(\ell_2)}\right| \;\leqslant\; q^{2-\ell_1-\ell_2} N^{(1/3-2k)+C\varepsilon} p^{-6+2k}\,. \tag{6.73}$$

Inserting (6.71) and (6.73) into the second term of the left-hand side of (6.63), and using the assumption F as well as (6.62), we find

$$\begin{aligned}
&\mathbb{E}\left(F'(x^R)(x_0^S - x_0^R) + F'(x^R)(x_1^S - x_1^R) + \frac{1}{2}F''(x^R)(x_0^S - x_0^R)^2\right) \\
&\quad=\; B + \mathbb{E}F'(x^R) Q_0^{(3)} + O\left(N^{-1/3+C\varepsilon} p^{-4} q^{-1}\right) \\
&\quad=\; B + \mathbb{E}F'(x^R)\left(Q_0^{(0,3)} + Q_0^{(3,0)}\right) + O\left(N^{-1/3+C\varepsilon} p^{-4} q^{-1}\right)\,,
\end{aligned} \tag{6.74}$$

where we defined

$$
\begin{aligned}
B \;&:=\; \mathbb{E}\left(\sum_{k=0,1} F'(x^R)\Big(Q_k^{(1)}+Q_k^{(2)}\Big)+\frac{1}{2}F''(x^R)\Big(Q_0^{(1)}\Big)^2\right)\\
&=\; \mathbb{E}\left(\sum_{k=0,1} F'(x^R)\mathbb{E}_{bd}\Big(Q_k^{(1)}+Q_k^{(2)}\Big)+\frac{1}{2}F''(x^R)\mathbb{E}_{bd}\Big(Q_1^{(0)}\Big)^2\right)\,. \qquad (6.75)
\end{aligned}
$$

Note that B depends on h_{bd} only through its expectation (which is zero) and on its second moment. Thus, B will be $B_N(m_2(v_{bd}),\mathrm{law}(Q))$ from (6.51).

In order to estimate (6.74), it only remains to estimate $\mathbb{E}F'(x^R)Q_0^{(0,3)}$ and $\mathbb{E}F'(x^R)Q_0^{(3,0)}$. Using (6.70), (6.63), (6.74), and (6.71), we have

$$
\Big|\mathbb{E}\big(F(x^S)-F(x^R)\big)-B\Big| \;\leqslant\; N^{-1/3+C\varepsilon}p^{-3}q^{-1}\,, \qquad (6.76)
$$

which implies (6.54) in the case $b=d$.

Let us therefore from now on assume $b\neq d$. Since we estimate $Q_0^{(3,0)}$ and $Q_0^{(0,3)}$, this implies that i,j,b,d are all distinct. In order to enforce this condition in sums, it is convenient to introduce the indicator function $\chi\equiv\chi(i,j,b,d):=\mathbf{1}\big(|\{i,j,b,d\}| \;=\; 4\big)$.

Recalling (6.64), we introduce the notation $R_{ij}^{(m,s)}$ to denote the sum of the terms in the definition (6.64) of $R_{ij}^{(m)}$ in which the number of the off-diagonal elements of R is s. For example,

$$
R_{ij}^{(3,0)}=R_{ij}^{(3,1)} \;=\; 0\,, \qquad R_{ij}^{(3,2)} \;=\; R_{ib}h_{bd}R_{dd}h_{db}R_{bb}h_{bd}R_{dj}+R_{id}h_{db}R_{bb}h_{bd}R_{dd}h_{db}R_{bj}\,. \qquad (6.77)
$$

Then in the case $\chi=1$ we have

$$
R_{ij}^{(3)}=\sum_{s=2}^{4}R_{ij}^{(3,s)} \qquad (6.78)
$$

Now from the definition (6.66) we get

$$
Q_0^{(0,3)} \;=\; \sum_{s=2}^{4}Q_0^{(0,3,s)} \qquad \text{where} \quad Q_0^{(0,3,s)} \;:=\; \eta\int_{E_1}^{E_2}\mathrm{d}y\sum_{i,j}\chi\,R_{ij}^{(0)}\overline{R_{ji}^{(3,s)}}\,, \qquad (6.79)
$$

and

$$
Q_0^{(3,0)} \;=\; \sum_{s=2}^{4}Q_0^{(3,0,s)} \qquad \text{where} \quad Q_0^{(3,0,s)} \;:=\; \eta\int_{E_1}^{E_2}\mathrm{d}y\sum_{i,j}\chi\,R_{ij}^{(3,s)}\overline{R_{ji}^{(0)}}\,. \qquad (6.80)
$$

As above, it is easy to see that for $s\geqslant 3$ we have

$$
\mathbb{E}F'(x^R)Q_0^{(0,3,s)} \;\leqslant\; N^{-1/3+C\varepsilon}p^{-4}q^{-1}\,, \qquad (6.81)
$$

which implies, using (6.74),

$$
\big|\mathbb{E}\big(F(x^S)-F(x^R)\big)-B\big| \;\leqslant\; \mathbb{E}F'(x^R)Q_0^{(0,3,2)}+\mathbb{E}F'(x^R)Q_0^{(3,0,2)}+N^{-5/3+C\varepsilon}q^{-1}\,. \qquad (6.82)
$$

By symmetry, it only remains to prove that

$$\mathbb{E}F'(x^R)Q_0^{(3,0,2)} \;\leqslant\; N^{-1/3+C\varepsilon}p^{-4}q^{-1}\,. \tag{6.83}$$

Using the definition (6.79) and the estimate (6.44) to replace some diagonal resolvent matrix elements with m_{sc}, we find

$$\begin{aligned}\mathbb{E}F'(x^R)Q_0^{(3,0,2)} \;&=\; \eta\int_{E_1}^{E_2}\mathrm{d}y\,\mathbb{E}F'(x^R)\sum_{i,j}\chi\left[R_{ib}h_{bd}R_{dd}h_{db}R_{bb}h_{bd}R_{dj}\overline{R_{ji}}+(b\leftrightarrow d)\right]\\ &=\; \eta\int_{E_1}^{E_2}\mathrm{d}y\,\mathbb{E}F'(x^R)\sum_{i,j}\chi\left[m_{sc}^2R_{ib}R_{dj}\overline{R_{ji}}\left(\mathbb{E}_{bd}|h_{bd}|^2h_{bd}\right)(b\leftrightarrow d)\right]\\ &\quad+O(N^{-1/3+C\varepsilon}p^{-4}q^{-1})\,,\end{aligned} \tag{6.84}$$

where we used the estimate $\left|\mathbb{E}_{bd}|h_{bd}|^2h_{bd}\right|\leqslant\frac{C}{Nq}$ to control the errors for the replacement. Combining (6.84) with (6.74) and (6.63), we therefore obtain

$$\begin{aligned}\left|\mathbb{E}[F(x^S)-F(x^R)]-B\right| \;&\leqslant\; CN^{-1/3+C\varepsilon}p^{-4}q^{-1}+Cq^{-1}N^{-1/3+C\varepsilon}\\ &\quad\times\max_{y\in I_\varepsilon}\max_{i,j}\chi\left[\left|\mathbb{E}F'(x^R)R_{ij}\overline{R_{jb}R_{di}}\right|+\left|\mathbb{E}F'(x^R)R_{ib}R_{dj}\overline{R_{ji}}\right|+(b\leftrightarrow d)\right],\end{aligned} \tag{6.85}$$

where we used the trivial bounds on F' and $|m_{sc}|$, and that every estimate is uniform in y.

In order to complete the proof of Lemma 6.7, we need to estimate the expectations in (6.85) by a better bound than the naive high-probability bound on the argument of $\mathbb{E}$. This is accomplished in Lemma 6.8 below. From Lemma 6.8 and (6.85) we get in the case $b\neq d$ that

$$\left|\mathbb{E}[F(x^S)-F(x^R)]-B\right| \;\leqslant\; N^{-1/3+C\varepsilon}p^{-4}q^{-1}\,, \tag{6.86}$$

where B was defined in (6.75). This completes the proof of Lemma 6.7. □

Lemma 6.8. *Under the assumptions of Lemma 6.7, in particular fixing $f=f_*$, and assuming that a,b,i,j are all distinct, we have*

$$\max_{y\in I_\varepsilon}\left|\mathbb{E}F'(x^R)R_{ib}R_{dj}\overline{R_{ji}}(y+\mathrm{i}\eta)\right| \;\leqslant\; Cp^{-4}\,. \tag{6.87}$$

The same estimate holds for the other three terms on the right-hand side of (6.85).

In order to prove Lemma 6.8, we shall need the following result, which is crucial when estimating terms arising from the nonvanishing expectation of $\mathbb{E}a_{ij}=fN^{-1}$. Before stating it, we introduce some notation.

Recall that we set $A\equiv A_{\gamma-1}=(a_{ij})$, where the matrix entries are given by $a_{ij}=h_{ij}+f/N$ and $\mathbb{E}h_{ij}=0$. We denote by $A^{(b)}$ the matrix obtained from A by setting all entries with index b to zero, i.e. $(A^{(b)})_{ij}:=\mathbf{1}(i\neq b)\mathbf{1}(j\neq b)a_{ij}$. If $Z\equiv Z(A)$ is a function of A, we define $Z^{(b)}:=Z(A^{(b)})$. See also Definitions 5.2 and 3.3 in [11]. We also use the notation $\mathbb{E}_b$ to denote partial expectation with respect to all variables $(a_{1b},\dots,a_{Nb})$ in the b-th column of A.

Lemma 6.9. *For any fixed i we have, with (ξ,ν)-high probability,*

$$\left|\frac{1}{N}\sum_{k\neq i}\sum_{l\neq k}S_{il}^{(k)}h_{lk}\right| \;\leqslant\; p^{-2}\,.$$

Proof. The claim is an immediate consequence of Proposition 7.11 in [11] and the observation that, for $E \in I_\varepsilon$, $\eta = N^{-2/3-\varepsilon}$, and $q \geqslant N^\phi$ we have

$$(\log N)^C \left(\frac{1}{q} + \sqrt{\frac{\operatorname{Im} m_{sc}}{N\eta}} + \frac{1}{N\eta} \right) \;\leqslant\; p^{-1}$$

for large enough N. □

Another ingredient necessary for the proof of Lemma 6.8 is the following resolvent identity.

Lemma 6.10. *Let $A = (a_{ij})$ be a square matrix and set $S = (S_{ij}) = (A - z)^{-1}$. Then for $i \neq j$ we have*

$$S_{ij} \;=\; -S_{ii} \sum_{k \neq i} a_{ik} S_{kj}^{(i)}\,, \qquad S_{ij} \;=\; -S_{jj} \sum_{k \neq j} S_{ik}^{(j)} a_{kj}\,. \tag{6.88}$$

Proof. We prove the first identity in (6.88); the second one is proved analogously. We use the resolvent identity

$$S_{ij} \;=\; S_{ij}^{(k)} + \frac{S_{ik} S_{kj}}{S_{kk}} \qquad \text{for } i, j \neq k \tag{6.89}$$

from [11], (3.8). Without loss of generality we assume that $z = 0$. Then (6.89) and the identity $AS = \mathbb{1}$ yield

$$\sum_{k \neq i} a_{ik} S_{kj}^{(i)} \;=\; \sum_{k \neq i} a_{ik} S_{kj} - \sum_{k \neq i} a_{ik} \frac{S_{ki} S_{ij}}{S_{ii}} \;=\; -a_{ii} S_{ij} - \frac{S_{ij}}{S_{ii}} (1 - a_{ii} S_{ii}) \;=\; \frac{S_{ij}}{S_{ii}}\,.$$ □

Armed with Lemmas 6.9 and 6.10, we may now prove Lemma 6.8.

Proof of Lemma 6.8. With the relation between R and S in (6.45) and (6.59), we find that (6.87) is implied by

$$\left| \mathbb{E} F'(x^S) S_{ib} S_{dj} \overline{S_{ji}} \right| \;\leqslant\; C p^{-4} N^{C\varepsilon}, \tag{6.90}$$

under the assumption that b, d, i, j are all distinct. This replacement is only a technical convenience when we apply a large deviation estimate below.

Recalling the definition of Ω after (6.55), we get using (6.89)

$$|S_{ij} - S_{ij}^{(b)}| \;=\; \left| S_{ib} S_{bj} (S_{bb})^{-1} \right| \;\leqslant\; C p^{-2} \qquad \text{in } \Omega\,. \tag{6.91}$$

This yields

$$|x^S - (x^S)^{(b)}| \;\leqslant\; p^{-1} N^{C\varepsilon} \qquad \text{in } \Omega\,. \tag{6.92}$$

Similarly, we have

$$\left| S_{ib} S_{dj} \overline{S_{ji}} - S_{ib} S_{dj}^{(b)} \overline{S_{ji}^{(b)}} \right| \;\leqslant\; C p^{-4} \qquad \text{in } \Omega\,. \tag{6.93}$$

Hence by assumption on F we have

$$|\mathbb{E} F'(x^S) S_{ib} S_{dj} \overline{S_{ji}}| \;\leqslant\; \left| \mathbb{E} \Big(F'\big((x^S)^{(b)}\big) \Big) S_{ib} S_{dj}^{(b)} \overline{S_{ji}^{(b)}} \right| + O\left(p^{-4} N^{C\varepsilon}\right). \tag{6.94}$$

Since $(x^S)^{(b)}$ and $S_{dj}^{(b)} \overline{S_{ji}^{(b)}}$ are independent of the b-th row of A, we find from (6.94) that (6.90), and hence (6.87), is proved if we can show that

$$\mathbb{E}_b S_{ib} \;=\; O(p^{-2}) \tag{6.95}$$

for any fixed i and b.

What remains therefore is to prove (6.95). Using (6.55) and (6.91) we find in Ω that

$$S_{bb} \;=\; m_{sc} + O(p^{-1})\,, \qquad S_{ik}^{(b)} \;=\; O(p^{-1})\,. \tag{6.96}$$

Using $a_{kb} = h_{kb} + f/N$ we write

$$S_{ib} \;=\; -m_{sc} \sum_{k \neq b} S_{ik}^{(b)} \biggl(h_{kb} + \frac{f}{N} \biggr) - (S_{bb} - m_{sc}) \sum_{k \neq b} S_{ik}^{(b)} \biggl(h_{kb} + \frac{f}{N} \biggr)\,. \tag{6.97}$$

By (6.96) and the large deviation estimate (3.15) in [11], the second sum in (6.97) is bounded by $O(p^{-1})$ with (ξ, ν)-high probability. Therefore, using (6.96) and $\mathbb{E}_b h_{kb} = 0$, we get

$$\mathbb{E}_b S_{ib} \;=\; \frac{-m_{sc} f}{N} \sum_{k \neq b} S_{ik}^{(b)} + O(p^{-2}) \;=\; \frac{-m_{sc} f}{N} \sum_{k \neq i} S_{ik} + O(p^{-2})\,, \tag{6.98}$$

where in the second step we used (6.89).

In order to estimate the right-hand side of (6.98), we introduce the quantity

$$X \;:=\; \frac{1}{N} \sum_{k \neq i} \mathbb{E}_k S_{ik}\,.$$

Note that X depends on the index i, which is omitted from the notation as it is fixed. Using (6.88), (6.96), and (6.89) as above, we find with (ξ, ν)-high probability

$$\begin{aligned}
X \;&=\; \frac{-m_{sc}}{N} \sum_{k \neq i} \sum_{l \neq k} \mathbb{E}_k\, S_{il}^{(k)} \biggl(h_{lk} + \frac{f}{N} \biggr) + O(p^{-2}) \\
&=\; \frac{-m_{sc} f}{N^2} \sum_{k \neq i} \sum_{l \neq k} S_{il}^{(k)} + O(p^{-2}) \\
&=\; \frac{-m_{sc} f}{N} \sum_{l \neq i} S_{il} + O(p^{-2}) \\
&=\; -m_{sc} f\, X + O\biggl(\frac{1}{N} \sum_{l \neq i} \bigl(S_{il} - \mathbb{E}_l S_{il} \bigr) \biggr) + O(p^{-2})\,.
\end{aligned}$$

Now recall that the spectral parameter $z = E + \mathrm{i}\eta$ satisfies $E \in I_\varepsilon$ (see (6.27)) and $\eta = N^{-2/3-\varepsilon}$. Therefore (3.6) implies that $m_{sc}(z) = -1 + o(1)$. Recalling that $f = 1 + \varepsilon_0$, we therefore get, with (ξ, ν)-high probability,

$$X \;=\; O\biggl(\frac{1}{N} \sum_{l \neq i} \bigl(S_{il} - \mathbb{E}_l S_{il} \bigr) \biggr) + O(p^{-2})\,. \tag{6.99}$$

We now return to (6.98), and estimate, with (ξ, ν)-high probability

$$\mathbb{E}_b S_{ib} \;=\; \frac{-m_{sc} f}{N} \sum_{k \neq i} S_{ik} + O(q^{-2}) \;=\; -m_{sc} f X + O\biggl(\frac{1}{N} \sum_{k \neq i} \bigl(S_{ik} - \mathbb{E}_k S_{ik} \bigr) \biggr) + O(p^{-2})\,.$$

Together with (6.99) this yields, with (ξ,ν)-high probability,

$$\mathbb{E}_b S_{ib} \;=\; O\bigg(\frac{1}{N}\sum_{k\neq i}\big(S_{ik}-\mathbb{E}_k S_{ik}\big)\bigg)+O(p^{-2})\,. \tag{6.100}$$

In order to estimate the quantity in parentheses, we abbreviate $\mathbb{IE}_k Z := Z-\mathbb{E}_k Z$ for any random variable Z and write, using (6.88),

$$\begin{aligned}\frac{1}{N}\sum_{k\neq i}\big(S_{ik}-\mathbb{E}_k S_{ik}\big) \;&=\; \frac{-1}{N}\sum_{k\neq i}\sum_{l\neq k}\mathbb{IE}_k\, S_{kk}S_{il}^{(k)}a_{lk}\\ &=\; \frac{-m_{sc}}{N}\sum_{k\neq i}\sum_{l\neq k}\mathbb{IE}_k\, S_{il}^{(k)}\bigg(h_{lk}+\frac{f}{N}\bigg)-\frac{1}{N}\sum_{k\neq i}\mathbb{IE}_k(S_{kk}-m_{sc})\sum_{l\neq k}S_{il}^{(k)}\bigg(h_{lk}+\frac{f}{N}\bigg)\,.\end{aligned}$$

Using the large deviation estimate (3.15) in [11], (6.96), and the bound $|h_{lk}|\leqslant p^{-1}$ which holds with (ξ,ν)-high probability (see Lemma 3.7 in [11]), we find that the second term is bounded by $O(p^{-2})$ with (ξ,ν)-high probability. Thus we get

$$\frac{1}{N}\sum_{k\neq i}\big(S_{ik}-\mathbb{E}_k S_{ik}\big) \;=\; \frac{-m_{sc}}{N}\sum_{k\neq i}\sum_{l\neq k}S_{il}^{(k)}h_{lk}+O(p^{-2})$$

with (ξ,ν)-high probability. Therefore (6.100) and Lemma 6.9 imply (6.95), and the proof is complete. □

7. Universality of generalized Wigner matrices with finite moments

This section is an application of our results to the problem of universality of generalized Wigner matrices (see Definition 7.1 below) whose entries have heavy tails. We prove the bulk universality of generalized Wigner matrices under the assumption that the matrix entries have a finite m-th moment for some $m>4$. We also prove the edge universality of Wigner matrices under the assumption that $m>12$. (This lower bound can in fact be improved to $m\geqslant 7$; see Remark 7.5 below.) The Tracy-Widom law for the largest eigenvalue of Wigner matrices was first proved in [33] under a Gaussian decay assumption, and was proved later in [29, 35, 19, 24] under various weaker restrictions on the distributions of the matrix elements. In particular, in [24] the Tracy-Widom law was proved for entries with symmetric distribution and $m>12$. In [23] similar results were derived for complex Hermitian Gaussian divisible matrices, where the GUE component is of order one. For this case it is proved in [23] that bulk universality holds provided the entries of the Wigner component have finite second moments, and edge universality holds provided they have finite fourth moments.

Definition 7.1. *We call a Hermitian or real symmetric random matrix $H=(h_{ij})$ a* generalized Wigner matrix *if the two following conditions hold. First, the family of upper-triangluar entries $(h_{ij}: i\leqslant j)$ is independent. Second, we have*

$$\mathbb{E}h_{ij} \;=\; 0\,,\qquad \mathbb{E}|h_{ij}|^2 \;=\; \sigma_{ij}^2\,,$$

where the variances σ_{ij}^2 satisfy

$$\sum_j \sigma_{ij}^2 \;=\; 1\,,\qquad C_-\leqslant \inf_{i,j}(N\sigma_{ij}^2)\;\leqslant\;\sup_{i,j}(N\sigma_{ij}^2)\;\leqslant\; C_+\,,$$

and $0 < C_- \leqslant C_+ < \infty$ are constants independent of N.

THEOREM 7.2 (BULK UNIVERSALITY). *Suppose that $H = (h_{ij})$ satisfies Definition 7.1. Let $m > 4$ and assume that for all i and j we have*

$$\mathbb{E}\big|h_{ij}/\sigma_{ij}\big|^m \;\leqslant\; C_m\,, \tag{7.1}$$

for some constant C_m, independent of i, j, and N.

Let $n \in \mathbb{N}$ and $O : \mathbb{R}^n \to \mathbb{R}$ be compactly supported and continuous. Let E satisfy $-2 < E < 2$ and let $\varepsilon > 0$. Then for any sequence b_N satisfying $N^{-1+\varepsilon} \leqslant b_N \leqslant ||E|-2|/2$ we have

$$\lim_{N\to\infty}\int_{E-b_N}^{E+b_N}\frac{\mathrm{d}E'}{2b_N}\int \mathrm{d}\alpha_1\cdots\mathrm{d}\alpha_n\, O(\alpha_1,\dots,\alpha_n) \\ \times\frac{1}{\varrho_{sc}(E)^n}\big(p_N^{(n)}-p_{\mathrm{G},N}^{(n)}\big)\bigg(E'+\frac{\alpha_1}{N\varrho_{sc}(E)},\dots,E'+\frac{\alpha_n}{N\varrho_{sc}(E)}\bigg) \;=\; 0\,.$$

Here ϱ_{sc} was defined in (2.10), $p_N^{(n)}$ is the n-point marginal of the eigenvalue distribution of H, and $p_{\mathrm{G},N}^{(n)}$ the n-point marginal of an $N \times N$ GUE/GOE matrix.

THEOREM 7.3 (EDGE UNIVERSALITY). *Suppose that $H^{\mathbf{v}} = (h_{ij}^{\mathbf{v}})$ and $H^{\mathbf{w}} = (h_{ij}^{\mathbf{w}})$ both satisfy Definition 7.1. Assume that the entries of $H^{\mathbf{v}}$ and $H^{\mathbf{w}}$ all satisfy (7.1) for some $m > 12$, and that the two first two moments of the entries of $h_{ij}^{\mathbf{v}}$ and $h_{ij}^{\mathbf{w}}$ match:*

$$\mathbb{E}^{\mathbf{v}}(\overline{h_{ij}^{\mathbf{v}}})^l(h_{ij}^{\mathbf{v}})^u \;=\; \mathbb{E}^{\mathbf{w}}(\overline{h_{ij}^{\mathbf{w}}})^l(h_{ij}^{\mathbf{w}})^u \qquad \text{for} \quad 0 \leqslant l+u \leqslant 2\,.$$

Then there is a $\delta > 0$ such that for any $s \in \mathbb{R}$ we have

$$\mathbb{P}^{\mathbf{v}}\Big(N^{2/3}(\lambda_N-2)\leqslant s-N^{-\delta}\Big)-N^{-\delta} \;\leqslant\; \mathbb{P}^{\mathbf{w}}\big(N^{2/3}(\lambda_N-2)\leqslant s\big) \;\leqslant\; \mathbb{P}^{\mathbf{v}}\Big(N^{2/3}(\lambda_N-2)\leqslant s+N^{-\delta}\Big)+N^{-\delta}\,. \tag{7.2}$$

Here $\mathbb{P}^{\mathbf{v}}$ and $\mathbb{P}^{\mathbf{w}}$ denote the laws of the ensembles $H^{\mathbf{v}}$ and $H^{\mathbf{w}}$ respectively, and λ_N denotes the largest eigenvalue of $H^{\mathbf{v}}$ or $H^{\mathbf{w}}$.

REMARK 7.4. A similar result holds for the smallest eigenvalue λ_1. Moreover, a result analogous to (7.2) holds for the n-point joint distribution functions of the extreme eigenvalues. (See [19], Equation (2.40)).

REMARK 7.5. With some additional effort, one may in fact improve the condition $m > 12$ in Theorem 7.3 to $m \geqslant 7$. The basic idea is to match seven instead of four moments in Lemma 7.7, and to use the resolvent expansion method from Section 6.3. We omit further details.

The rest of this section is devoted to the proof of Theorems 7.2 and 7.3.

7.1. Truncation. For definiteness, we focus on real symmetric matrices, but the following truncation argument applies trivially to complex Hermitian matrices by truncating the real and imaginary parts separately. To simplify the presentation, we consider Wigner matrices for which $\sigma_{ij} = N^{-1/2}$. The proof for the more general matrices from Definition 7.1 is the same; see also Remark 2.4 in [11].

We begin by noting that, without loss of generality, we may assume that the distributions of the entries of H are absolutely continuous. Otherwise consider the matrix $H + \varepsilon_N V$, where V is a GUE/GOE matrix independent of H, and (ε_N) is a positive sequence that tends to zero arbitrarily fast. (Note that the following argument is insensitive to the size of ε_N.)

Let $H \equiv H^{\mathbf{x}} = (h^{\mathbf{x}}_{ij})$ be a Wigner matrix whose entries are of the form $h^{\mathbf{x}}_{ij} = N^{-1/2} x_{ij}$ for some x_{ij}. We assume that the family $(x_{ij} : i \leqslant j)$ is independent, and that each x_{ij} satisfies

$$\mathbb{E} x_{ij} \;=\; 0\,, \qquad \mathbb{E}|x_{ij}|^2 = 1\,.$$

Moreover, we assume that there is an $m > 4$ and a constant $C_m \geqslant 1$, independent of i, j, and N, such that

$$\mathbb{E}|x_{ij}|^m \;\leqslant\; C_m\,.$$

In a first step we construct a truncated Wigner matrix $H^{\mathbf{y}}$ from $H^{\mathbf{x}}$. This truncation is performed in the following lemma.

Lemma 7.6. *Fix $m > 2$ and let X be a real random variable, with absolutely continuous law, satisfying*

$$\mathbb{E}X \;=\; 0\,, \qquad \mathbb{E}X^2 \;=\; 1\,, \qquad \mathbb{E}|X|^m \;\leqslant\; C_m\,.$$

Let $\lambda > 0$. Then there exists a real random variable Y that satisfies

$$\mathbb{E}Y \;=\; 0\,, \qquad \mathbb{E}Y^2 \;=\; 1\,, \qquad |Y| \;\leqslant\; \lambda\,, \qquad \mathbb{P}(X \neq Y) \;\leqslant\; 2 C_m \lambda^{-m}\,.$$

Proof. We introduce the abbreviations

$$P \;:=\; \mathbb{P}(|X| > \lambda)\,, \qquad E \;:=\; \mathbb{E}\big(X\,\mathbf{1}(|X| > \lambda)\big)\,, \qquad V \;:=\; \mathbb{E}\big(X^2\,\mathbf{1}(|X| > \lambda)\big)\,.$$

Using the assumption on X, Markov's inequality, and Hölder's inequality, we find

$$P \;\leqslant\; C_m \lambda^{-m}\,, \qquad |E| \;\leqslant\; C_m \lambda^{-m+1}\,, \qquad V \;\leqslant\; C_m \lambda^{-m+2}\,. \tag{7.3}$$

The idea behind the construction of Y is to cut out the tail $|X| > \lambda$, to add appropriate Dirac weights at $\pm\lambda$, and to adjust the total probability by cutting out the values $X \in [-a, a]$, where a is an appropriately chosen small nonnegative number. For any t satisfying $0 \leqslant t \leqslant 1/2$, we choose a nonnegative number a_t such that $\mathbb{P}(X \in [-a_t, a_t]) = t$. Note that since X is absolutely continuous such a number a_t exists and the map $t \to a_t$ is continuous. Moreover, using $\mathbb{E}X^2 = 1$ and Markov's inequality we find that $a_t \leqslant 2$ for $t \leqslant 1/2$. We define the quantities

$$e_t \;:=\; \mathbb{E}\big(X \mathbf{1}(-a_t \leqslant X \leqslant a_t)\big)\,, \qquad v_t \;:=\; \mathbb{E}\big(X^2 \mathbf{1}(-a_t \leqslant X \leqslant a_t)\big)\,,$$

which satisfy the trivial bounds

$$|e_t| \;\leqslant\; 2t\,, \qquad v_t \;\leqslant\; 4t\,. \tag{7.4}$$

We shall remove the values $(-\infty, -\lambda) \cup [-a_t, a_t] \cup (\lambda, \infty)$ from the range of X, and replace them with Dirac weights at λ and $-\lambda$ with respective probabilities p and q. Thus we are led to the system

$$p + q \;=\; P + t\,, \qquad p - q \;=\; \lambda^{-1}(E + e_t)\,, \qquad p + q \;=\; \lambda^{-2}(V + v_t)\,. \tag{7.5}$$

In order to solve (7.5), we abbreviate the right-hand sides of the equations in (7.5) by $\alpha(t)$, $\beta(t)$, and $\gamma(t)$ respectively.

In a first step, we solve t from the equation $\alpha(t) = \gamma(t)$. To that end, we observe that $\alpha(0) \leqslant \gamma(0)$, as follows from the trivial inequality $V \geqslant \lambda^2 P$. Moreover, $\alpha(1/2) \gg \gamma(1/2)$, by (7.3) and (7.4). Since $\alpha(t)$

and $\gamma(t)$ are continuous, the equation $\alpha(t) = \gamma(t)$ has a solution t_0. Moreover, (7.3) and (7.4) imply that $t_0 \leqslant C_m \lambda^{-m} + 4\lambda^{-2} t_0$, from which we get that $t_0 \leqslant 2C_m \lambda^{-m}$. For the following we fix $t := t_0$.

In a second step, we solve the equations $p + q = \alpha(t)$ and $p - q = \beta(t)$ to get

$$p \;=\; \frac{\alpha(t) + \beta(t)}{2}\,, \qquad q \;=\; \frac{\alpha(t) - \beta(t)}{2}\,.$$

We now claim that $|\beta(t)| \leqslant \alpha(t)$. Indeed, a simple application of Cauchy-Schwarz yields $|\beta(t)| \leqslant (\alpha(t) + \gamma(t))/2 = \alpha(t)$. Hence p and q are nonnegative. Moreover, the bounds (7.3) and (7.4) yield

$$p + q \;\leqslant\; 2C_m \lambda^{-m}\,.$$

Thus we have proved that (7.5) has a solution (p, q, t) satisfying

$$0 \;\leqslant\; p, q, t \;\leqslant\; 2C_m \lambda^{-m}\,.$$

Next, let $I := (-\infty, -\lambda) \cup [-a_t, a_t] \cup (\lambda, \infty)$. Thus, $\mathbb{P}(X \in I) = p + q$. Partition $I = I_1 \cup I_2$ such that $\mathbb{P}(X \in I_1) = p$ and $\mathbb{P}(X \in I_2) = q$. Then we define

$$Y \;:=\; X \mathbf{1}(X \notin I) + \lambda \mathbf{1}(X \in I_1) - \lambda \mathbf{1}(X \in I_2)\,.$$

Recalling (7.5), we find that Y satisfies $\mathbb{E}Y = 0$ and $\mathbb{E}Y^2 = 1$. Moreover,

$$\mathbb{P}(X \neq Y) \;=\; \mathbb{P}(X \in I) \;=\; p + q \;\leqslant\; 2C_m \lambda^{-m}\,.$$

This concludes the proof. $\square$

Note that the variable Y constructed in Lemma 7.6 satisfies $\mathbb{E}|Y|^m \leqslant 3C_m$. Let ρ be some exponent satisfying $0 < \rho < 1/2$ and assume that $m > 4$. Using Lemma 7.6 with $\lambda := N^\rho$, we construct, for each x_{ij}, a random variable y_{ij} such that the family $(y_{ij} : i \leqslant j)$ is independent and

$$\mathbb{E}y_{ij} \;=\; 0\,, \qquad \mathbb{E}y_{ij}^2 \;=\; 1\,, \qquad |y_{ij}| \;\leqslant\; N^\rho\,, \qquad \mathbb{P}(x_{ij} \neq y_{ij}) \;\leqslant\; 2C_m N^{-\rho m}\,, \qquad \mathbb{E}|y_{ij}|^m \;\leqslant\; 3C_m\,. \tag{7.6}$$

We define the new matrix $H^{\mathbf{y}} = (h^{\mathbf{y}}_{ij})$ through $h^{\mathbf{y}}_{ij} := N^{-1/2} y_{ij}$. In particular, we have

$$\mathbb{E}h^{\mathbf{y}}_{ij} \;=\; 0\,, \qquad \mathbb{E}|h^{\mathbf{y}}_{ij}|^2 \;=\; \frac{1}{N}\,, \qquad \mathbb{E}|h^{\mathbf{y}}_{ij}|^p \;\leqslant\; \frac{1}{N q^{p-2}}\,, \tag{7.7}$$

where we set

$$q \;:=\; N^{1/2 - \rho}\,. \tag{7.8}$$

Thus, $H^{\mathbf{y}}$ satisfies Definition 2.1.

7.2. Moment matching. Next, we construct a third Wigner matrix, $H^{\mathbf{z}} = (h^{\mathbf{z}}_{ij})$, whose entries are of the form $h^{\mathbf{z}}_{ij} = N^{-1/2} z_{ij}$. We require that z_{ij} have uniformly subexponential decay, i.e.

$$\mathbb{E}z_{ij} \;=\; 0\,, \qquad \mathbb{E}|z_{ij}|^2 \;=\; 1\,, \qquad \mathbb{P}(|z_{ij}| \geqslant \xi) \;\leqslant\; \theta^{-1} \mathrm{e}^{-\xi^\theta}\,, \tag{7.9}$$

for some $\theta > 0$ independent of i, j, and N. We choose z_{ij} so as to match the first four moments of y_{ij}.

LEMMA 7.7. *Let y_{ij} satisfy* (7.6) *with some $m > 4$. Then there exists a z_{ij} satisfying* (7.9) *such that $\mathbb{E}z_{ij}^l = \mathbb{E}y_{ij}^l$ for $l = 1, \ldots, 4$.*

PROOF. In fact, using an explicit construction similar to the one used in the proof of Theorem 2.5, z_{ij} can be chosen to be supported at only three points. We omit further details. □

It was proved in [19], Section 2.1, that the statement of Theorem 7.2 holds if the entries of H satisfy the subexponential decay condition (7.9). Theorem 7.2 will therefore follow if we can prove that, for b_N and O as in Theorem 7.2, we have

$$\lim_{N\to\infty} \int_{E-b_N}^{E+b_N} \frac{\mathrm{d}E'}{2b_N} \int \mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_n \, O(\alpha_1, \ldots, \alpha_n) \\ \times \frac{1}{\varrho_{sc}(E)^n} \big(p_{\mathbf{x},N}^{(n)} - p_{\mathbf{z},N}^{(n)}\big) \left(E' + \frac{\alpha_1}{N\varrho_{sc}(E)}, \ldots, E' + \frac{\alpha_n}{N\varrho_{sc}(E)}\right) \;=\; 0\,, \tag{7.10}$$

where $p_{\mathbf{x},N}^{(n)}$ and $p_{\mathbf{z},N}^{(n)}$ are the n-point marginals of the eigenvalue distributions of $H^{\mathbf{x}}$ and $H^{\mathbf{z}}$, respectively.

Similarly, it was proved in [19], Section 2.2, that the statement of Theorem 7.3 holds if the entries of $H^{\mathbf{v}}$ and $H^{\mathbf{w}}$ both satisfy the subexponential decay condition (7.9). Thus, Theorem 7.3 will follow if we can prove that

$$\mathbb{P}^{\mathbf{x}}\Big(N^{2/3}(\lambda_N - 2) \leqslant s - N^{-\delta}\Big) - N^{-\delta} \;\leqslant\; \mathbb{P}^{\mathbf{z}}\big(N^{2/3}(\lambda_N - 2) \leqslant s\big) \;\leqslant\; \mathbb{P}^{\mathbf{x}}\Big(N^{2/3}(\lambda_N - 2) \leqslant s + N^{-\delta}\Big) + N^{-\delta}\,, \tag{7.11}$$

for some $\delta > 0$. Here we use $\mathbb{P}^{\mathbf{x}}$ and $\mathbb{P}^{\mathbf{z}}$ to denote the laws of ensembles $H^{\mathbf{x}}$ and $H^{\mathbf{z}}$ respectively.

We shall prove both (7.10) and (7.11) by first comparing $H^{\mathbf{x}}$ to $H^{\mathbf{y}}$ and then comparing $H^{\mathbf{y}}$ to $H^{\mathbf{z}}$. The first step is easy: from (7.6) we get

$$\mathbb{P}(H^{\mathbf{x}} \neq H^{\mathbf{y}}) \;\leqslant\; 2C_m N^{2-\rho m}\,. \tag{7.12}$$

Thus, (7.10) and (7.11) hold with $\mathbf{z}$ replaced by $\mathbf{y}$ provided that

$$\rho m \;>\; 2\,. \tag{7.13}$$

7.3. Comparison of $H^{\mathbf{y}}$ and $H^{\mathbf{z}}$ in the bulk. In this section we prove that (7.10) holds with $\mathbf{x}$ replaced by $\mathbf{y}$, and hence complete the proof of Theorem 7.3.

We compare the local spectral statistics of $H^{\mathbf{y}}$ and $H^{\mathbf{z}}$ using the Green function comparison method from [17], Section 8. The key additional input is the local semicircle theorem for sparse matrices, Theorem 3.3. We merely sketch the differences to [17]. As explained in [17], the n-point correlation functions $p_N^{(n)}$ can be expressed (up to an error N^{-c}) in terms of expectations of observables F, whose arguments are products of expressions of the form $m(z_i + \mathrm{i}\eta)$ where $\eta := N^{-1-\varepsilon}$. We assume that the first five derivatives of F are polynomially bounded, uniformly in N. Using the local semicircle law for sparse matrices, Theorem 3.3, we may control the Green function matrix elements down to scales $N^{-1-\varepsilon}$, uniformly in E. (Note that in [17], the spectral edge had to be excluded since the bounds derived there were unstable near the edge, unlike our bound (3.14).) This allows us to compare the local eigenvalue statistics of the matrix ensembles at scales $N^{-1-\varepsilon}$, which is sufficiently accurate for both Theorems 7.2 and 7.3.

We use the telescopic summation and the Lindeberg replacement argument from [17], Chapter 8, whose notations we take over without further comment; see also Section 6.3. A resolvent expansion yields

$$S \;=\; R - N^{-1/2}RVR + N^{-1}(RV)^2R - N^{-3/2}(RV)^3R + N^{-2}(RV)^4R - N^{-5/2}(RV)^5S\,.$$

Next, note that, by (7.7) and (7.9), both ensembles $H^{\mathbf{y}}$ and $H^{\mathbf{z}}$ satisfy Definition 2.1 with q defined in (7.8). Hence we may invoke Theorem 3.3 with $f = 0$, and in particular (3.14), for the matrices $H^{\mathbf{y}}$ and $H^{\mathbf{z}}$. Choosing $\eta = N^{-1-\varepsilon}$ for some $\varepsilon > 0$, we therefore get

$$|R_{ij}(E + \mathrm{i}\eta)| \leqslant N^{2\varepsilon}, \qquad |S_{ij}(E + \mathrm{i}\eta)| \leqslant N^{2\varepsilon}$$

with (ξ, ν)-high probability.

Consider now the difference $\mathbb{E}^{\mathbf{y}} - \mathbb{E}^{\mathbf{z}}$ applied to some fixed observable F depending on traces (normalized by N^{-1}) of resolvents and whose derivatives have at most polynomial growth. Since the first four moments of the entries of $H^{\mathbf{y}}$ and $H^{\mathbf{z}}$ coincide by Lemma 7.7, the error in one step of the telescopic summation is bounded by the expectation of the rest term in the resolvent expansion, i.e.

$$N^{C\varepsilon} N^{-5/2} \mathbb{E}(RV)^5 S \leqslant N^{-5/2+C\varepsilon} \max_{a,b} \mathbb{E}|V_{ab}|^5 \leqslant N^{-5/2+C\varepsilon} C_m N^{\rho},$$

where in the last step we used (7.6). The first factor $N^{C\varepsilon}$ comes from the polynomially bounded derivatives of F. Summing up all $O(N^2)$ terms of the telescopic sum, we find that the difference $\mathbb{E}^{\mathbf{y}} - \mathbb{E}^{\mathbf{z}}$ applied to F is bounded by

$$N^{-1/2+C\varepsilon+\rho}. \tag{7.14}$$

Combining (7.14) and (7.12), we find that both (7.10) follows provided that

$$-\frac{1}{2} + C\varepsilon + \rho < 0, \qquad \rho m > 2. \tag{7.15}$$

Since $m > 4$ is fixed, choosing first $1/2 - \rho$ small enough and then ε small enough yields (7.15). This completes the proof of Theorem 7.2.

7.4. Comparison of $H^{\mathbf{y}}$ and $H^{\mathbf{z}}$ at the edge. In order to prove (7.2) under the assumption $m > 12$, we may invoke Proposition 6.4, which implies that (7.11) holds with $\mathbf{x}$ replaced by $\mathbf{y}$, provided that $\phi = 1/2 - \rho > 1/3$, i.e. $\rho < 1/6$. Together with the condition (7.13), this implies that (7.2) holds if $m > 12$.

A. Regularization of the Dyson Brownian motion

In this appendix we sketch a simple regularization argument needed to prove two results concerning the Dyson Brownian motion (DBM). This argument can be used as a substitute for earlier, more involved, proofs given in Appendices A and B of [16] on the existence of the dynamics restricted to the subdomain $\Sigma_N := \{\mathbf{x} : x_1 < x_2 < \cdots < x_N\}$, and on the applicability of the Bakry-Emery method. The argument presented in this section is also more probabilistic in nature than the earlier proofs of [16].

For applications in Section 4 of this paper, some minor adjustments to the argument below are needed to incorporate the separate treatment of the largest eigenvalue. These modifications are straightforward, and we shall only sketch the argument for the standard DBM.

Theorem A.1. *Fix $n \geqslant 1$ and let $\mathbf{m} = (m_1, \dots, m_n) \in \mathbb{N}^n$ be an increasing family of indices. Let $G : \mathbb{R}^n \to \mathbb{R}$ be a continuous function of compact support and set*

$$\mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) := G\big(N(x_i - x_{i+m_1}), N(x_{i+m_1} - x_{i+m_2}), \dots, N(x_{i+m_{n-1}} - x_{i+m_n})\big).$$

Let $\gamma_1, \ldots, \gamma_N$ denote the classical locations of the eigenvalues and set

$$Q \;:=\; \sup_{t \in [t_0, \tau]} \sum_{i=1}^{N} \int (x_i - \gamma_i)^2 f_t \,\mathrm{d}\mu \tag{A.1}$$

Choose an $\varepsilon > 0$. Then for any ρ satisfying $0 < \rho < 1$, and setting $\tau = N^{-\rho}$, there exists a $\bar{\tau} \in [\tau/2, \tau]$ such that, for any $J \subset \{1, 2, \ldots, N - m_n - 1\}$, we have

$$\biggl|\int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, f_{\bar{\tau}} \,\mathrm{d}\mu - \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \,\mathrm{d}\mu\biggr| \;\leqslant\; C N^{\varepsilon} \sqrt{\frac{NQ + 1}{|J| \tau}} \tag{A.2}$$

for all $N \geqslant N_0(\rho)$. Here $\mu = \mu^{(N)}$ is the equilibrium measure of the N eigenvalues of the GOE.

Define $\mu_\beta(\mathbf{x}) = C e^{-N\beta\mathcal{H}(\mathbf{x})}$ as in (4.6) and (4.5), but introducing a parameter β so that μ_β is the equilibrium measure of the usual β-ensemble which is invariant under the (β-dependent) DBM. We remark that Theorem A.1 holds for all $\beta \geqslant 1$ with an identical proof. The following lemma holds more generally for $\beta > 0$.

Let $\omega := C \mu_\beta e^{-N \sum_j U_j(x_j)}$, where U_j is a C^2-function satisfying

$$\min_j U_j''(x) \;\geqslant\; \tau^{-1} \tag{A.3}$$

for some $\tau < 1$. For the following lemma we recall the definition (4.7) of the Dirichlet form.

Lemma A.2. *Let $\beta > 0$ and $q \in H^1(\mathrm{d}\omega)$ be a probability density with respect to ω. Then for any $\beta > 0$ and any $J \subset \{1, 2, \ldots, N - m_n - 1\}$ and any $t > 0$ we have*

$$\biggl|\int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \, q \,\mathrm{d}\omega - \int \frac{1}{|J|} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}} \,\mathrm{d}\omega\biggr| \;\leqslant\; C \sqrt{\frac{D_\omega(\sqrt{q})\, t}{|J|}} + C \sqrt{S_\omega(q)}\, \mathrm{e}^{-ct/\tau}\,. \tag{A.4}$$

Recall that the DBM is defined via the stochastic differential equation

$$\mathrm{d}x_i \;=\; \frac{\mathrm{d}B_i}{\sqrt{N}} + \beta\biggl(-\frac{1}{4} x_i + \frac{1}{2N} \sum_{j \neq i} \frac{1}{x_i - x_j}\biggr)\mathrm{d}t \qquad \text{for} \quad i = 1, \ldots, N\,, \tag{A.5}$$

where $B_1, \ldots, B_N$ is a family of independent standard Brownian motions. It was proved in [1], Lemma 4.3.3, that there is a unique strong solution to (A.5) for all $\beta \geqslant 1$.

For any $\delta > 0$ define the extension μ_β^δ of the measure μ_β from Σ_N to $\mathbb{R}^N$ by replacing the singular logarithm with a C^2-function. To that end, we introduce the approximation parameter $\delta > 0$ and define, as in Section 4, for $\mathbf{x} \in \mathbb{R}^N$,

$$\mathcal{H}_\delta(\mathbf{x}) \;:=\; \sum_i \frac{1}{4} x_i^2 - \frac{1}{N} \sum_{i<j} \log_\delta(x_j - x_i)$$

where we set

$$\log_\delta(x) \;:=\; \mathbf{1}(x \geqslant \delta) \log x + \mathbf{1}(x < \delta) \biggl(\log \delta + \frac{x - \delta}{\delta} - \frac{1}{2\delta^2}(x - \delta)^2\biggr)\,.$$

It is easy to check that $\log_\delta \in C^2(\mathbb{R})$, is concave, and satisfies

$$\lim_{\delta\to 0} \log_\delta(x) \;=\; \begin{cases} \log x & \text{if } x > 0 \\ -\infty & \text{if } x \leqslant 0\,. \end{cases}$$

Furthermore, we have the lower bound

$$\partial_x^2 \log_\delta(x) \;\geqslant\; \begin{cases} -\frac{1}{x^2} & \text{if } x > \delta \\ -\frac{1}{\delta^2} & \text{if } x \leqslant \delta\,. \end{cases}$$

Similarly, we can extend the measure ω to $\mathbb{R}^N$ by setting $\omega^\delta = Ce^{-N\sum_j U_j(x_j)}\mu_\beta^\delta$.

Lemma A.3. *Let $q \in L^\infty(\mathrm{d}\omega^\delta)$ be a C^2 probability density. Then for $\delta \leqslant 1/N$, $\beta > 0$ and any $J \subset \{1, 2, \ldots, N - m_n - 1\}$ and any $t > 0$ we have*

$$\left|\int \frac{1}{|J|}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}}\, q \,\mathrm{d}\omega^\delta - \int \frac{1}{|J|}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}} \,\mathrm{d}\omega^\delta\right| \;\leqslant\; C\sqrt{\frac{D_{\omega^\delta}(\sqrt{q})\,t}{|J|}} + C\sqrt{S_{\omega^\delta}(q)}\,\mathrm{e}^{-ct/\tau}\,. \tag{A.6}$$

Proof. The proof of Theorem 4.3 in [16] applies with merely cosmetic changes; now however the dynamics is defined on $\mathbb{R}^N$ instead of Σ_N, so that complications arising from the boundary are absent. The condition $\delta \leqslant 1/N$ is needed since we use the singularity of $\partial_x^2 \log x$ to generate a factor $1/N^2$ in the regime $x \leqslant C/N$ in the proof. □

Proof of Lemma A.2. Suppose that q is a probability density in Σ_N with respect to ω. We extend q to be zero outside Σ and let $q_\varepsilon \in C^2$ be any regularization of q on $\mathbb{R}^N$ that converges to q in $H^1(\omega)$. Then there is a constant $C_{\varepsilon,\delta}$ such that $q_\varepsilon^\delta := C_{\varepsilon,\delta}\, q_\varepsilon$ is a probability density with respect to ω_δ. Thus (A.6) holds with q replaced by q_ε^δ. Taking the limit $\delta \to 0$ and then $\varepsilon \to 0$, we have

$$\begin{aligned} &\left|\int \frac{1}{|J|}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}}\, q \,\mathrm{d}\omega - \int \frac{1}{|J|}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}} \,\mathrm{d}\omega\right| \\ &\qquad\qquad \leqslant\; C\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\sqrt{\frac{D_{\omega^\delta}(\sqrt{q_\varepsilon^\delta})\,t}{|J|}} + C\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\sqrt{S_{\omega^\delta}(q_\varepsilon^\delta)}\,\mathrm{e}^{-ct/\tau}\,. \end{aligned} \tag{A.7}$$

Notice that $\omega_\delta \to \mathbf{1}(\Sigma_N)\omega$ weakly as $\delta \to 0$. Thus

$$\lim_{\varepsilon\to 0}\lim_{\delta\to 0} D_{\omega^\delta}(\sqrt{q_\varepsilon^\delta}) \;=\; \lim_{\varepsilon\to 0} D_\omega(\sqrt{q_\varepsilon}) \;=\; D_\omega(\sqrt{q})$$

provided that $q \in H^1(\omega)$. This proves Lemma A.2. Notice that the proof did not use the existence of DBM; instead, it used the existence of the regularized DBM. □

Proof of Theorem A.1. Write

$$\int \frac{1}{|J|}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}}\, f_t \,\mathrm{d}\mu \;=\; \mathbb{E}^{f_0\mu}\mathbb{E}^{\mathbf{x}_0}\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}(\mathbf{x}(t))\,.$$

Here $\mathbb{E}^{\mathbf{x}_0}$ denotes expectation with respect to the law of the DBM $(\mathbf{x}(t))_t$ starting from $\mathbf{x}_0$, and $\mathbb{E}^{f_0\mu}$ denotes expectation of $\mathbf{x}_0$ with respect to the measure $f_0\mu$. Let $\mathbb{E}_\delta$ denote expectation with respect to the regularized DBM. Then we have

$$\mathbb{E}^{\mathbf{x}_0}\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}(\mathbf{x}(t)) \;=\; \lim_{\delta\to 0}\mathbb{E}_\delta^{\mathbf{x}_0}\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}(\mathbf{x}(t)),$$

where we have used the existence of a strong solution to the DBM (see [1], Lemma 4.3.3) and that the dynamics remains in Σ_N almost surely. Hence

$$\int\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}\, f_t\,\mathrm{d}\mu \;=\; \lim_{\delta\to 0}\int\frac{1}{|J|}\sum_{i\in J}\mathcal{G}_{i,\mathbf{m}}\, f_t^\delta\,\mathrm{d}\mu^\delta,$$

where f_t^δ is the solution to the regularized DBM at the time t with initial data $f_0\mu/\mu^\delta$. Using that (A.2) holds for the regularized dynamics, and taking the limit $\delta\to 0$, we complete the proof. □

References

[1] Anderson, G., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices. Studies in advanced mathematics, **118**, Cambridge University Press, 2009.

[2] Aptekarev, A., Khabibullin, R.: Asymptotic expansions for polynomials orthogonal with respect to a complex non-constant weight function, *Trans. Moscow Math. Soc.* **68**, 1–37 (2007).

[3] Auffinger, A., Ben Arous, G., Péché, S.: Poisson Convergence for the largest eigenvalues of heavy-tailed matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **45**, no. 3, 589–610 (2009).

[4] Biroli, G., Bouchaud, J.-P., Potters, M.: On the top eigenvalue of heavy-tailed random matrices. *Europhysics Lett.* **78**, 10001 (2007).

[5] Bleher, P., Its, A.: Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model. *Ann. of Math.* **150**, 185–266 (1999).

[6] Brascamp, H. J. and Lieb, E. H.: On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22**, 366–389 (1976).

[7] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.* **52**, 1335–1425 (1999).

[8] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: Strong asymptotics of orthogonal polynomials with respect to exponential weights. *Comm. Pure Appl. Math.* **52**, 1491–1552 (1999).

[9] Dyson, F.J.: A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.* **3**, 1191–1198 (1962).

[10] Erdős, L.: Universality of Wigner random matrices: a Survey of Recent Results (lecture notes). Preprint arXiv:1004.0861.

[11] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of sparse random matrices I: local semicircle law. Preprint arXiv:1103.1919, to appear in *Ann. Prob.*

[12] Erdős, L., Schlein, B., Yau, H.-T.: Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37**, No. 3, 815–852 (2009).

[13] Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for Wigner random matrices. *Commun. Math. Phys.* **287**, 641–655 (2009).

[14] Erdős, L., Schlein, B., Yau, H.-T.: Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Notices.* **2010**, No. 3, 436-479 (2010).

[15] Erdős, L., Schlein, B., Yau, H.-T.: Universality of random matrices and local relaxation flow. Preprint arXiv:0907.5605.

[16] Erdős, L., Schlein, B., Yau, H.-T., Yin, J.: The local relaxation flow approach to universality of the local statistics for random matrices. To appear in *Ann. Inst. H. Poincaré Probab. Statist.* Preprint arXiv:0911.3687.

[17] Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices. Preprint arXiv:1001.3453.

[18] Erdős, L., Yau, H.-T., Yin, J.: Universality for generalized Wigner matrices with Bernoulli distribution. To appear in *J. Combinatorics.* Preprint arXiv:1003.3813.

[19] Erdős, L., Yau, H.-T., Yin, J.: Rigidity of eigenvalues of generalized Wigner matrices. Preprint arXiv:1007.4652.

[20] Erdős, P.; Rényi, A.: On random graphs. I. *Publicationes Mathematicae* **6**, 290–297 (1959).

[21] Erdős, P.; Rényi, A.: The evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**: 17–61 (1960).

[22] Johansson, K.: Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.* **215**, no. 3., 683–705 (2001).

[23] Johansson, K.: Universality for certain Hermitian Wigner matrices under weak moment conditions. *Ann. Inst. H. Poincaré Probab. Statist.* **48**, 47–79 (2012).

[24] Khorunzhi, O.: High moments of large Wigner random matrices and asymptotic properties of the spectral norm. To appear in *Rand. Op. Stoch. Eqs.*

[25] Knowles, A., Yin, J.: Eigenvector distribution of Wigner matrices. Preprint arXiv:1102.0057, to appear in *Prob. Theor. Rel. Fields.*

[26] Miller, S. J., Novikoff, T., Sabelli, A.: The distribution of the largest nontrivial eigenvalues in families of random regular graphs. *Exper. Math.* **17**, 231–244 (2008).

[27] Pastur, L., Shcherbina M.: Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles. *J. Stat. Phys.* **86**, 109–147 (1997).

[28] Pastur, L., Shcherbina M.: Bulk universality and related properties of Hermitian matrix models. *J. Stat. Phys.* **130**, 205–250 (2008).

[29] Ruzmaikina, A.: Universality of the edge distribution of eigenvalues of Wigner random matrices with polynomially decaying distributions of entries, *Comm. Math. Phys.* **261**, no. 2, 277–296 (2006).

[30] Sarnak, P.: Private communication.

[31] Sinai, Y. and Soshnikov, A.: A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge. *Functional Anal. and Appl.* **32**, no. 2, 114–131 (1998).

[32] Sodin, S.: The Tracy–Widom law for some sparse random matrices. Preprint arXiv:0903.4295.

[33] Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207**, no. 3, 697–733 (1999).

[34] Tao, T. and Vu, V.: Random matrices: Universality of the local eigenvalue statistics, to appear in *Acta Math.*, Preprint arXiv:0906.0510.

[35] Tao, T. and Vu, V.: Random matrices: Universality of local eigenvalue statistics up to the edge. Preprint arXiv:0908.1982.

[36] Tracy, C., Widom, H.: Level-Spacing Distributions and the Airy Kernel, *Comm. Math. Phys.* **159**, 151–174 (1994).

[37] Tracy, C., Widom, H.: On orthogonal and symplectic matrix ensembles, *Comm. Math. Phys.* **177**, no. 3, 727–754 (1996).